(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 December 2005 (22.12.2005)

PCT

(10) International Publication Number
WO 2005/121331 A2

(51) International Patent Classification:
C12N 9/10 (2006.01)

(21) International Application Number:
PCT/US2005/019442

(22) International Filing Date: 3 June 2005 (03.06.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/576,530	3 June 2004 (03.06.2004)	US
60/598,584	3 August 2004 (03.08.2004)	US

(71) **Applicant** *(for all designated States except US)*: **NEOSE TECHNOLOGIES, INC.** [US/US]; 102 Witmer Road, Horsham, PA 19044 (US).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: **JOHNSON, Karl, F.** [US/US]; 2941 Windsor Ave., Willow Grove, PA 19090 (US). **CHEN, Xi** [CN/US]; 997 Hunter Lane, Woodland, CA (US). **TAUDTE, Susann** [DE/US]; Pennsburg, PA (US). **SARIBAS, Sami** [TR/US]; 424 East Church Lane, Philadelphia, PA 19144 (US).

(74) **Agents: KELLY, Beth L.** et al.; Townsend and Townsend and Crew LLP, 2 Embarcadero Center, 8th Floor, San Francisco, CA 94111 (US).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

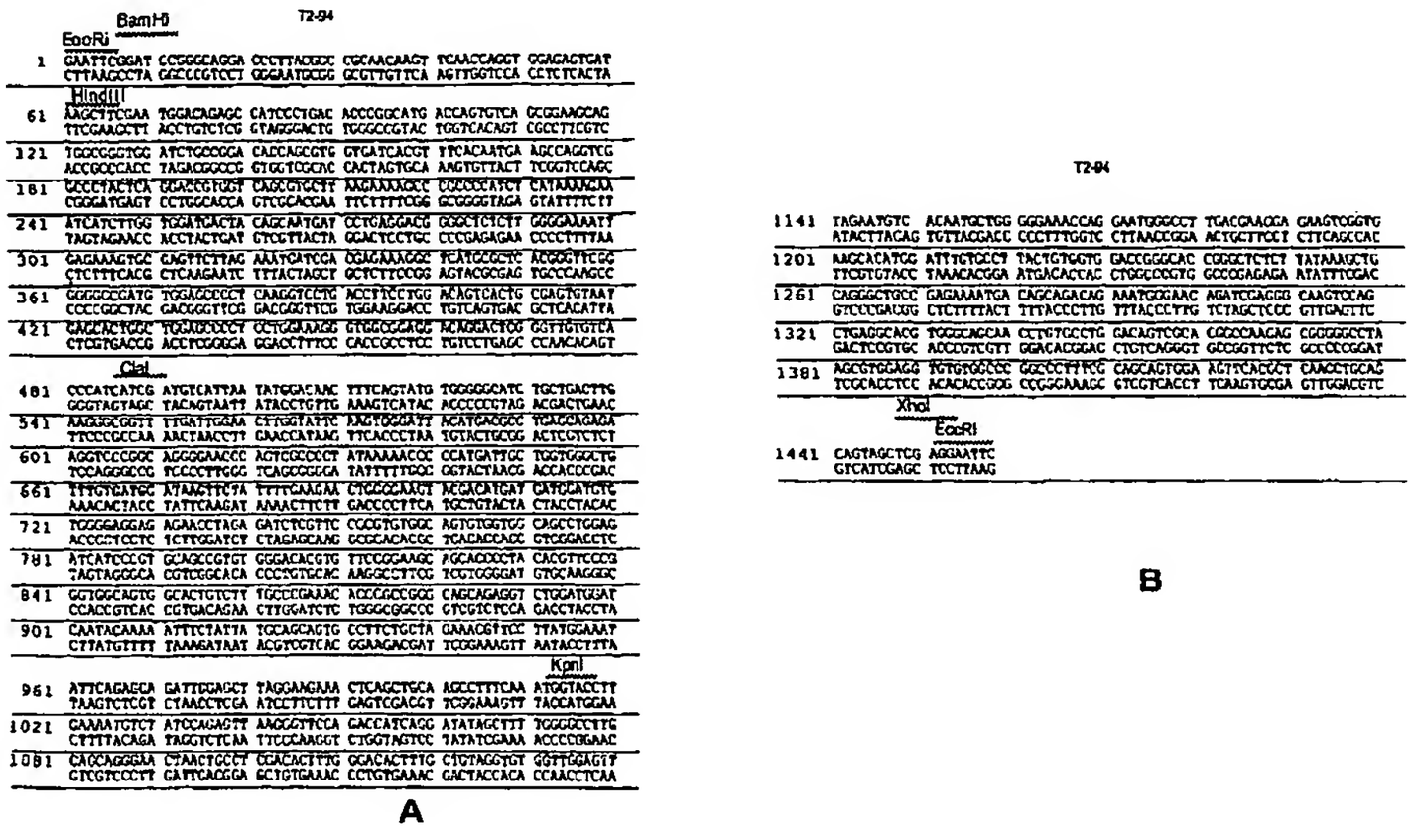
Published:

— *without international search report and to be republished upon receipt of that report*

(48) **Date of publication of this corrected version:**
1 June 2006

[Continued on next page]

(54) **Title:** TRUNCATED GALNACT2 POLYPEPTIDES AND NUCLEIC ACIDS



(57) **Abstract:** The present invention features compositions and methods related to truncated mutants of GalNAcT2. In particular, the invention features truncated human GalNAcT2 polypeptides. The invention also features nucleic acids encoding such truncated polypeptides, as well as vectors, host cells, expression systems, and methods of expressing and using such polypeptides.

WO 2005/121331 A2

(15) **Information about Corrections:**
see PCT Gazette No. 22/2006 of 1 June 2006
Previous Correction:
see PCT Gazette No. 10/2006 of 9 March 2006

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TRUNCATED GALNACT2 POLYPEPTIDES AND NUCLEIC ACIDS

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 60/576,530, filed June 3, 2004 and U.S. Provisional Application No. 60/598,584, filed August 3, 2004; both of which are herein incorporated by reference for all purposes.

FIELD OF THE INVENTION

[0002] The present invention features compositions and methods related to truncated mutants of GalNAcT2. In particular, the invention features truncated human GalNAcT2 polypeptides. The invention also features nucleic acids encoding such truncated polypeptides, as well as vectors, host cells, expression systems, and methods of expressing and using such polypeptides.

BACKGROUND OF THE INVENTION

[0003] A great diversity of oligosaccharide structures and many types of glycopeptides are found in nature, and these are synthesized, in part, by a large number of glycosyltransferases. Glycosyltransferases catalyze the synthesis of glycolipids, glycopeptides, and polysaccharides, by transferring an activated mono- or oligosaccharide residue to an existing acceptor molecule for the initiation or elongation of the carbohydrate chain. A catalytic reaction is believed to involve the recognition of both the donor and acceptor by suitable domains, as well as the catalytic site of the enzyme.

[0004] Many peptide therapeutics, and many potential peptide therapeutics, are glycosylated peptides. The production of a recombinant glycopeptide, as opposed to a recombinant non-glycosylated peptide, requires that a recombinantly-produced peptide is subjected to additional processing steps, either within the cell or after the peptide is produced by the cell, where the processing steps are performed in vitro. The peptide can be treated enzymatically to introduce one or more glycosyl groups onto the peptide, using a glycosyltransferase. Specifically, the glycosyltransferase covalently attaches the glycosyl group or groups to the peptide.

[0005] The extra in vitro steps of peptide processing to produce a glycopeptide can be time consuming and costly. This is due, in part, to the burden and cost of producing recombinant glycosyltransferases for the in vitro glycosylation of peptides and glycopeptides to produce

glycopeptide therapeutics. As the demand and usefulness of recombinant glycotherapeutics increases, new methods are required in order to more efficiently prepare glycopeptides. Moreover, as more and mo4re glycopeptides are discovered to be useful for the treatment of a variety of diseases, there is a need for methods that lower the cost of their production. Further, there is also a need in the art to develop methods of more efficiently producing recombinant glycopeptides for use in developing and improving glycopeptide therapeutics.

[0006] Glycosyltransferases are reviewed in general in International (PCT) Patent Application No. WO03/031464 (PCT/US02/32263), which is incorporated herein by reference in its entirety. One such particular glycosyltransferase that has utility in the development and production of therapeutic glycopeptides is GalNAcT2. GalNAcT2, or N-acetyl-D-galactosamine transferase, catalyzes the transfer of GalNAc from a GalNAc donor to a GalNAc acceptor. Full length human GalNAcT2 enzyme is disclosed by Bennett et al. (1996, J Biol Chem. 271:17006-17012). However, the identification of useful mutants of this enzyme, having enhanced biological activity such as enhanced catalytic activity or enhanced stability, has not heretofore been reported.

[0007] In the past, there have been efforts to increase the availability of recombinant glycosyltransferases for the in vitro production of glycopeptides. A limited amount of work has been done with respect to recombinant glycosyltransferases that may sometimes be suitable for small-scale production of oligosaccharides or glycopeptides. For example, White et al. have disclosed a soluble form of human GalNAcT2 (1995, J. Biol. Chem., 270:24156-24165). Additionally, Kurosawa et al. (1994, J Biol Chem. 269:1402-1409) describe a truncation mutant of chicken GalNAcα2,6-sialyltransferase (ST6GalNAcI) lacking amino acid residues 1-232 from the full-length enzyme. However, the truncated enzyme described by Kurosawa et al. lacks the substrate specificity of other ST6GalNAcI enzymes. Therefore, a need still exists for recombinant glycosyltransferases having activity that is suitable for "pharmaceutical-scale" processes and reactions, including the production of glycopeptide therapeutics. In particular, there is a need for recombinant glycosyltranasferases having favorable functional and structural characteristics. Further, a need exists for efficient methods of identification and characterization of recombinant glycosyltransferases, as well as for the production of such glycosyltransferases. The present invention addresses and meets these needs.

BRIEF SUMMARY OF THE INVENTION

[0008] In one aspect, the present invention provides an isolated nucleic acid comprising a nucleic acid sequence that encodes a truncated human GalNAcT2 polypeptide. The truncated human GalNAcT2 polypeptide lacks all or a portion of the GalNAcT2 signal domain, or in addition lacks all or a portion the GalNAcT2 transmembrane domain, or in addition lacks all or a portion the GalNAcT2 stem domain; with the proviso that the encoded polypeptide is not a human GalNAcT2 truncation mutant polypeptide lacking amino acid residues 1-51.

[0009] In one embodiment, the isolated nucleic acid comprises a nucleic acid sequence having at least 90% identity with a nucleic acid selected from the group consisting of SEQ ID NO:3, SEQ ID NO:7 and SEQ ID NO:9. In another embodiment, the isolated nucleic acid comprises a nucleic acid sequence having at least 95% identity with a nucleic acid selected from the group consisting of SEQ ID NO:3, SEQ ID NO:7 and SEQ ID NO:9. In a further embodiment, the isolated nucleic acid comprises a nucleic acid sequence selected from SEQ ID NO:3, SEQ ID NO:7 and SEQ ID NO:9.

[0010] In some embodiments, the isolated nucleic acid is an isolated chimeric nucleic acid encoding a fusion polypeptide. The fusion polypeptide can include a tag polypeptide covalently linked to a truncated human GalNAcT2 polypeptide, as described herein. Examples of tag polypeptides include a maltose binding protein, a histidine tag, a Factor IX tag, a glutathione-S-transferase tag, a FLAG-tag, and a starch binding domain tag.

[0011] In another aspect, the invention provides an isolated truncated human GalNAcT2 polypeptide, that lacks all or a portion of the GalNAcT2 signal domain, or in addition lacks all or a portion the GalNAcT2 transmembrane domain, or in addition lacks all or a portion the GalNAcT2 stem domain; with the proviso that the encoded polypeptide is not a human GalNAcT2 truncation mutant polypeptide lacking amino acid residues 1-51. IN one embodiment, the isolated truncated human GalNAcT2 polypeptide has at least 90% or 95% identity with a polypeptide selected from the group consisting of SEQ ID NO:4, SEQ ID NO:8 and SEQ ID NO:10. In a further aspect, isolated truncated human GalNAcT2 polypeptide comprises an amino acid sequence selected from SEQ ID NO:4, SEQ ID NO:8 and SEQ ID NO:10.

[0012] In some embodiments, the isolated truncated GalNAcT2 polypeptide isolated chimeric polypeptide comprising a tag polypeptide covalently linked to the isolated truncated GalNAcT2. Examples of tag polypeptides include a maltose binding protein, a histidine tag,

a Factor IX tag, a glutathione-S-transferase tag, a FLAG-tag, and a starch binding domain tag.

[0013] The isolated nucleic acid encoding a truncated GalNAcT2 polypeptide can also be operably linked to a promoter/regulatory sequence, within *e.g.*, an expression vector. The invention also includes host cells that comprise such expression vectors. Host cells can be *e.g.*, eukaryotic or a prokaryotic cells. Eukaryotic cells include, *e.g.*, mammalian cells, an insect cells, and a fungal cells. Some preferred mammalian host cells are SF9 cells, an SF9+ cells, an Sf21 cells, a HIGH FIVE cells or Drosophila Schneider S2 cells. Prokaryotic host cells include, *e.g.*, *E. coli* cells and *B. subtilis* cells.

[0014] The host cells can be used to producing a truncated human GalNAcT2 polypeptide, by growing the recombinant host cells of under conditions suitable for expression of the truncated human GalNAcT2 polypeptide. In preferred embodiments, sufficient truncated human GalNAcT2 polypeptide is made to allow commercial scale production of a glycoprotein or glycopeptide.

[0015] In a further aspect the invention includes a method of catalyzing the transfer of a GalNAc moiety to an acceptor moiety comprising incubating the truncated human GalNAcT2 polypeptide with a GalNAc moiety and an acceptor moiety, wherein said polypeptide mediates the covalent linkage of said GalNAc moiety to said acceptor moiety, thereby catalyzing the transfer of a GalNAc moiety to an acceptor moiety to produce a product saccharide, or a product glycoprotein, or a product glycopeptide. In one embodiment, the acceptor moiety is a granulocyte colony stimulating factor (G-CSF) protein. In another embodiment, the acceptor moiety is selected from erythropoietin, human growth hormone, granulocyte colony stimulating factor, interferons alpha, -beta, and –gamma, Factor IX, follicle stimulating hormone, interleukin-2, erythropoietin, anti-TNF-alpha, and a lysosomal hydrolase. In a further embodiment, the polypeptide acceptor is a glycopeptide. In some embodiments, the GalNAc moiety comprises a polyethylene glycol moiety. In another embodiment, the product saccharide, product glycoprotein, or product glycopeptide is produced on a commercial scale.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] For purpose of illustrating the invention, there are depicted in the drawings certain embodiments of the invention. However, the invention is not limited to the precise arrangements and instrumentalities of the embodiments depicted in the drawings.

[0017] Figure 1 is an image of an electrophoretic gel illustrating the PCR amplification of ppGalNAcT2 genes. M, 1 kb DNA ladder; PCR1, PCR product for ppGalNAcT2-N41R (1596 bp); PCR2, PCR product for ppGalNAcT2-N52K (1563 bp); PCR3, PCR product for ppGalNAcT2-N74G (1497 bp); PCR4, PCR product for ppGalNAcT2-N95G (1434 bp).

[0018] Figure 2A is a plasmid restriction map for the pCWin2MBP vector.

[0019] Figure 2B is an image of an electrophoretic gel illustrating the fragments resulting from multiple samples of the pCWin2MBP vector digested by both BamHI and XhoI restriction enzymes.

[0020] Figure 3 is an image of an electrophoretic gel illustrating the screening of DH5α (pCWin2MBP-ppGalNAcT2) colonies by restriction mapping (BamHI and XhoI digestion) for plasmid purified from twelve colonies. Lane M, bp ladder. Lanes 1-3, N41R; lanes 4-6, N52K; lanes 8-10, N74G; lanes 11-13, N95G.

[0021] Figure 4 is an image of an electrophoretic protein gel illustrating SDS-PAGE for JM109 (pCWin2MBP-ppGalNAcT2) whole cell lysates after IPTG induction as described elsewhere herein. M, Pre-Stained MW Standard; Lane 13, IPTG-induced JM109 (pCWin2MBP); Lanes1-12, protein in whole cells for colonies 1-12 ; Lanes1-3, JM109 (pCWin2MBP-ppGalNAcT2N41R); Lanes 4-6, JM109 (pCWin2MBP-ppGalNAcT2N52K); Lanes 7-9, JM109 (pCWin2MBP-ppGalNAcT2N74G); Lanes 10-12, JM109 (pCWin2MBP-ppGalNAcT2N95G).

[0022] Figure 5 is an image of an electrophoretic protein gel illustrating SDS-PAGE for JM109 (pCWin2MBP-ppGalNAcT2) cell lysates. M, Pre-Stained MW Standard; Lane 13, lysate from JM109 (pCWin2MBP); Lanes 1-12, lysates from colonies 1-12; Lanes 1-3, JM109 (pCWin2MBP-ppGalNAcT2N41R); Lanes 4-6, JM109 (pCWin2MBP-ppGalNAcT2N52K); Lanes 7-9, JM109 (pCWin2MBP-ppGalNAcT2N74G); Lanes 10-12, JM109 (pCWin2MBP-ppGalNAcT2N95G).

[0023] Figure 6 is an image of an electrophoretic protein gel illustrating SDS-PAGE for inclusion bodies isolated from JM109 (pCWin2MBP-ppGalNAcT2) cells. M, Pre-Stained MW Standard; Lane 13, inclusion bodies from JM109 (pCWin2MBP); Lanes 1-12, inclusion bodies from colonies 1-12; Lanes 1-3, JM109 (pCWin2MBP-ppGalNAcT2N41R); Lanes 4-6, JM109 (pCWin2MBP-ppGalNAcT2N52K); Lanes 7-9, JM109 (pCWin2MBP-ppGalNAcT2N74G); Lanes 10-12, JM109 (pCWin2MBP-ppGalNAcT2N95G).

[0024] Figure 7 is an image of an electrophoretic gel illustrating the protein expression pattern in lysates of cells containing human GalNAcT2 constructs. Lane 1, molecular weight marker; lane 2, construct 1 culture before induction; lane 3, construct 1 culture after induction; lane 4, construct 2 culture before induction; lane 5, construct 2 culture after induction; lane 6, construct 3 culture before induction; lane 7, construct 3 culture after induction; lane 8, construct 4 culture before induction; lane 9, construct 4 culture after induction; lane 10, empty.

[0025] Figure 8 is an image of an electrophoretic protein gel illustrating the protein content of inclusion bodies from JM109 pCWin2 MBP-GalNAcT2 constructs. Lane 1, MW marker; lane 2, JM109 pCWin2 MBP-GalNAcT2 construct 1 inclusion bodies; lane 3, JM109 pCWin2 MBP-GalNAcT2 construct 2 inclusion bodies.

[0026] Figure 9 is an image of an electrophoretic protein gel illustrating the glycoPEGylation of G-CSF by Δ51 GalNAcT2-MBP. Lane 1, glycoPEGylation in the presence of 1 mg/ml G-CSF; lane 2, glycoPEGylation in the presence of 0.7 mg/ml G-CSF; lane 3, glycoPEGylation in the presence of 0.4 mg/ml G-CSF; lane 4, glycoPEGylation in the presence of 0.2 mg/ml G-CSF. The glycoPEGylated G-CSF is visible around 60 kDa.

[0027] Figures 10A and 10B depict a nucleic acid sequence encoding a Δ40 GalNAcT2 polypeptide.

[0028] Figures 11A and 11B depict a nucleic acid sequence encoding a Δ51 GalNAcT2 polypeptide.

[0029] Figures 12A and 12B depict a nucleic acid sequence encoding a Δ73 GalNAcT2 polypeptide.

[0030] Figures 13A and 13B depict a nucleic acid sequence encoding a Δ94 GalNAcT2 polypeptide.

[0031] Figure 14A is an image of a chromatogram illustrating the elution of Δ51 GalNAcT2-MBP that was refolded at pH 5.5 and subsequently eluted from a Q-sepharose fast flow column. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis.

[0032] Figure 14B is an image of two electrophoretic gels used to visualize the eluted fractions set forth in Figure 14A. The contents of each lane on the gel are described in the

figure.Figure 14C is a table illustrating the relative GalNAc transferase activity of the fractions set forth in Figure 14A.

[0033] Figure 15A is an image of a chromatogram illustrating the elution of Δ51 GalNAcT2-MBP that was refolded at pH 6.5 and subsequently eluted from a Q-sepharose fast flow column. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis.

[0034] Figure 15B is an image of two electrophoretic gels used to visualize the eluted fractions set forth in Figure 15A. The contents of each lane on the gel are described in the figure.

[0035] Figure 15C is a table illustrating the relative GalNAc transferase activity of the fractions set forth in Figure 15A.

[0036] Figure 16A is an image of a chromatogram illustrating the elution of Δ51 GalNAcT2-MBP that was refolded at pH 8.0 and subsequently eluted from a Q-sepharose fast flow column. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis.

[0037] Figure 16B is an image of two electrophoretic gels used to visualize the eluted fractions set forth in Figure 16A. The contents of each lane on the gel are described in the figure.

[0038] Figure 16C is a table illustrating the relative GalNAc transferase activity of the fractions set forth in Figure 16A.

[0039] Figure 17A is an image of a chromatogram illustrating the elution of Δ51 GalNAcT2-MBP that was refolded at pH 8.5 and subsequently eluted from a Q-sepharose fast flow column. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis.

[0040] Figure 17B is an image of two electrophoretic gels used to visualize the eluted fractions set forth in Figure 17A. The contents of each lane on the gel are described in the figure.

[0041] Figure 17C is a table illustrating the relative GalNAc transferase activity of the fractions set forth in Figure 17A.

[0042] Figure 18A is an image of a chromatogram illustrating the elution of Δ51 GalNAcT2-MBP that was refolded at pH 8.0 and subsequently eluted from a Q-sepharose fast flow column. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis.

[0043] Figure 18B is an image of two electrophoretic gels used to visualize the eluted fractions set forth in Figure 18A. The contents of each lane on the gel are described in the figure.

[0044] Figure 18C is a table illustrating the relative GalNAc transferase activity of the fractions set forth in Figure 18A.

[0045] Figure 19A is an image of a chromatogram illustrating the elution of Δ51 GalNAcT2-MBP from a Q-sepharose fast flow column. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis.

[0046] Figure 19B is an image of two electrophoretic gels used to visualize the eluted fractions set forth in Figure 19A. The contents of each lane on the gel are described in the figure and correspond to the chromatogram of Figure 19A.

[0047] Figure 19C is a table illustrating the relative GalNAc transferase activity of the fractions set forth in Figure 19A.

[0048] Figure 20A is an image of a chromatogram illustrating the elution of Δ51 GalNAcT2-MBP from a Q-sepharose XL column, using 5 mM NaCl. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis.

[0049] Figure 20B is an image of two electrophoretic gels used to visualize the eluted fractions set forth in Figure 20A. The contents of each lane on the gel are described in the figure and correspond to the chromatogram of Figure 20A.

[0050] Figure 20C is a table illustrating the relative GalNAc transferase activity of the fractions set forth in Figure 20A.

[0051] Figure 21A is an image of a chromatogram illustrating the elution of Δ51 GalNAcT2-MBP from a Q-sepharose XL column, using 50 mM NaCl. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis.

[0052] Figure 21B is an image of two electrophoretic gels used to visualize the eluted fractions set forth in Figure 21A. The contents of each lane on the gel are described in the figure and correspond to the chromatogram of Figure 21A.

[0053] Figure 21C is a table illustrating the relative GalNAc transferase activity of the fractions set forth in Figure 21A.

[0054] Figure 22A is an image of a chromatogram illustrating the elution of Δ51 GalNAcT2-MBP from a Q-sepharose XL column, using 100 mM NaCl. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis.

[0055] Figure 22B is an image of two electrophoretic gels used to visualize the eluted fractions set forth in Figure 22A. The contents of each lane on the gel are described in the figure and correspond to the chromatogram of Figure 22A.

[0056] Figure 22C is a table illustrating the relative GalNAc transferase activity of the fractions set forth in Figure 22A.

[0057] Figure 23A is an image of a chromatogram illustrating the elution of Δ51 GalNAcT2-MBP from a Q-sepharose XL column, using 200 mM NaCl. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis.

[0058] Figure 23B is an image of two electrophoretic gels used to visualize the eluted fractions set forth in Figure 23A. The contents of each lane on the gel are described in the figure and correspond to the chromatogram of Figure 23A.

[0059] Figure 23C is a table illustrating the relative GalNAc transferase activity of the fractions set forth in Figure 23A.

[0060] Figure 24A is an image of a chromatogram illustrating the elution of Δ51 GalNAcT2-MBP from a Hydroxyapatite Type I column. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis.

[0061] Figure 24B is an image of an electrophoretic gel used to visualize the eluted fractions set forth in Figure 24A. The contents of each lane on the gel are described in the figure and correspond to the chromatogram of Figure 24A.

[0062] Figure 24C is a table illustrating the relative GalNAc transferase activity of the fractions set forth in Figure 24A.

[0063] Figure 25 is a graph illustrating the relative GalNAc transferase activity of various preparations of refolded Δ51 GalNAcT2-MBP. The refolding conditions of each preparation is indicated on the x-axis, and the relative GalNAc transferase activity is illustrated on the Y-axis.

[0064] Figure 26 is a graph illustrating the relative GalNAc transferase activity of various preparations of refolded Δ51 GalNAcT2-MBP. The refolding conditions of each preparation is indicated on the x-axis, and the relative GalNAc transferase activity is illustrated on the Y-axis.

[0065] Figure 27 is an image of three MALDI-TOF spectra demonstrating GalNAc transfer to GCSF mediated by Δ51 GalNAcT2-MBP that has been refolded and purified according to the present invention.

[0066] Figure 28 is an image of three MALDI-TOF spectra demonstrating GalNAc transfer to GCSF mediated by Δ51 GalNAcT2-MBP that has been refolded and purified according to the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0067] The compositions and methods of the present invention encompass truncation mutants of human GalNAcT2 polypeptides, isolated nucleic acids encoding these proteins, and methods of their use. GalNAcT2 polypeptides catalyze the transfer of a GalNAc from a GalNAc donor to a GalNAc acceptor.

[0068] The glycosyltransferase GalNAcT2 is an essential reagent for glycosylation of therapeutic glycopeptides. Additionally, GalNAcT2 is an important reagent for research and development of therapeutically important glycopeptides and oligosaccharide therapeutics. GalNAcT2 enzymes are typically isolated and purified from natural sources, or from tedious and costly in vitro and recombinant sources. The present invention provides compositions and methods relating to simplified and more cost-effective methods of production of GalNAcT2 enzymes. In particular, the present invention provides compositions and methods relating to truncated GalNAcT2 enzymes that have improved and useful properties in comparison to their full-length enzyme counterparts.

[0069] Truncated glycosyltransferase enzymes of the present invention are useful for in vivo and in vitro preparation of glycosylated peptides, as well as for the production of oligosaccharides containing the specific glycosyl residues that can be transferred by the truncated glycosyltransferase enzymes of the present invention. This is because it is shown for the first time herein that truncated forms of GalNAcT2 polypeptides possess biological activities comparable to, and in some instances, in excess of their full-length polypeptide counterparts. The present application also discloses that such truncation mutants not only possess biological activity, but also that the truncation mutants may have enhanced properties of solubility, stability and resistance to proteolytic degradation.

Definitions

[0070] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described herein.

[0071] Certain abbreviations are used herein as are common in the art, such as: "Ac" for acetyl; "Glc" for glucose; “Glc” for glucosamine; “GlcA for glucuronic acid; “IdoA” for iduronic acid; "GlcNAc" for N-acetylglucosamine; “NAN” or “sialic acid” or “SA” for N-acetyl neuraminic acid; “UDP” for uridine diphosphate; “CMP” for cytidine monophosphate.

[0072] As used herein, each of the following terms has the meaning associated with it in this section.

[0073] The articles "a" and "an" are used herein to refer to one or to more than one (i.e. to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

[0074] “Encoding” refers to the inherent property of specific sequences of nucleotides in a nucleic acid, such as a gene, a cDNA, or an mRNA, to serve as templates for synthesis of other polymers and macromolecules in biological processes having either a defined sequence of nucleotides (i.e., rRNA, tRNA and mRNA) or a defined sequence of amino acids and the biological properties resulting therefrom. Thus, a gene encodes a protein if transcription and translation of mRNA corresponding to that gene produces the protein in a cell or other biological system. Both the coding strand, the nucleotide sequence of which is identical to the mRNA sequence and is usually provided in sequence listings, and the non-coding strand,

used as the template for transcription of a gene or cDNA, can be referred to as encoding the protein or other product of that gene or cDNA.

[0075] A "coding region" of a gene consists of the nucleotide residues of the coding strand of the gene and the nucleotides of the non-coding strand of the gene which are homologous with or complementary to, respectively, the coding region of an mRNA molecule which is produced by transcription of the gene.

[0076] A "coding region" of an mRNA molecule also consists of the nucleotide residues of the mRNA molecule which are matched with an anticodon region of a transfer RNA molecule during translation of the mRNA molecule or which encode a stop codon. The coding region may thus include nucleotide residues corresponding to amino acid residues which are not present in the mature protein encoded by the mRNA molecule (*e.g.*, amino acid residues in a protein export signal sequence).

[0077] An "affinity tag" is a peptide or polypeptide that may be genetically or chemically fused to a second polypeptide for the purposes of purification, isolation, targeting, trafficking, or identification of the second polypeptide. The "genetic" attachment of an affinity tag to a second protein may be effected by cloning a nucleic acid encoding the affinity tag adjacent to a nucleic acid encoding a second protein in a nucleic acid vector.

[0078] As used herein, the term "glycosyltransferase," refers to any enzyme/protein that has the ability to transfer a donor sugar to an acceptor moiety.

[0079] A "sugar nucleotide-generating enzyme" is an enzyme that has the ability to produce a sugar nucleotide. Sugar nucleotides are known in the art, and include, but are not limited to, such moieties as UDP-Gal, UDP-GalNAc, and CMP-NAN.

[0080] An "isolated nucleic acid" refers to a nucleic acid segment or fragment which has been separated from sequences which flank it in a naturally occurring state, e.g., a DNA fragment which has been removed from the sequences which are normally adjacent to the fragment, e.g., the sequences adjacent to the fragment in a genome in which it naturally occurs. The term also applies to nucleic acids which have been substantially purified from other components which naturally accompany the nucleic acid, e.g., RNA or DNA or proteins, which naturally accompany it in the cell. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a

separate molecule (e.g, as a cDNA or a genomic or cDNA fragment produced by PCR or restriction enzyme digestion) independent of other sequences. It also includes a recombinant DNA which is part of a hybrid gene encoding additional polypeptide sequence.

[0081] In the context of the present invention, the following abbreviations for the commonly occurring nucleic acid bases are used. "A" refers to adenosine, "C" refers to cytidine, "G" refers to guanosine, "T" refers to thymidine, and "U" refers to uridine.

[0082] A "polynucleotide" means a single strand or parallel and anti-parallel strands of a nucleic acid. Thus, a polynucleotide may be either a single-stranded or a double-stranded nucleic acid.

[0083] The term "nucleic acid" typically refers to large polynucleotides. However, the terms “nucleic acid” and “polynucleotide” are used interchangeably herein.

[0084] The term "oligonucleotide" typically refers to short polynucleotides, generally no greater than about 50 nucleotides. It will be understood that when a nucleotide sequence is represented by a DNA sequence (i.e., A, T, G, C), this also includes an RNA sequence (i.e., A, U, G, C) in which “U” replaces "T.”

[0085] Conventional notation is used herein to describe nucleic acid sequences: the left-hand end of a single-stranded nucleic acid sequence is the 5’ end; the left-hand direction of a double-stranded nucleic acid sequence is referred to as the 5'-direction.

[0086] A first defined nucleic acid sequence is said to be “immediately adjacent to” a second defined nucleic acid sequence when, for example, the last nucleotide of the first nucleic acid sequence is chemically bonded to the first nucleotide of the second nucleic acid sequence through a phosphodiester bond. Conversely, a first defined nucleic acid sequence is also said to be “immediately adjacent to” a second defined nucleic acid sequence when, for example, the first nucleotide of the first nucleic acid sequence is chemically bonded to the last nucleotide of the second nucleic acid sequence through a phosphodiester bond.

[0087] A first defined polypeptide sequence is said to be “immediately adjacent to” a second defined polypeptide sequence when, for example, the last amino acid of the first polypeptide sequence is chemically bonded to the first amino acid of the second polypeptide sequence through a peptide bond. Conversely, a first defined polypeptide sequence is said to be “immediately adjacent to” a second defined polypeptide sequence when, for example, the

first amino acid of the first polypeptide sequence is chemically bonded to the last amino acid of the second polypeptide sequence through a peptide bond.

[0088] The direction of 5' to 3' addition of nucleotides to nascent RNA transcripts is referred to as the transcription direction. The DNA strand having the same sequence as an mRNA is referred to as the "coding strand"; sequences on the DNA strand which are located 5' to a reference point on the DNA are referred to as "upstream sequences"; sequences on the DNA strand which are 3' to a reference point on the DNA are referred to as "downstream sequences."

[0089] Unless otherwise specified, a "nucleotide sequence encoding an amino acid sequence" includes all nucleotide sequences that are degenerate versions of each other and that encode the same amino acid sequence. Nucleotide sequences that encode proteins and RNA may include introns.

[0090] "Homology" as used herein, refers to nucleotide sequence similarity between two regions of the same nucleic acid strand or between regions of two different nucleic acid strands. When a nucleotide residue position in both regions is occupied by the same nucleotide residue, then the regions are homologous at that position. A first region is homologous to a second region if at least one nucleotide residue position of each region is occupied by the same residue. Homology between two regions is expressed in terms of the proportion of nucleotide residue positions of the two regions that are occupied by the same nucleotide residue. By way of example, a region having the nucleotide sequence 5'-ATTGCC-3' and a region having the nucleotide sequence 5'-TATGGC-3' share 50% homology. Preferably, the first region comprises a first portion and the second region comprises a second portion, whereby, at least about 50%, and preferably at least about 75%, at least about 90%, or at least about 95% of the nucleotide residue positionss of each of the portions are occupied by the same nucleotide residue. More preferably, all nucleotide residue positions of each of the portions are occupied by the same nucleotide residue.

[0091] As used herein, "percent identity" is used synonymously with "homology." The determination of percent identity between two nucleotide or amino acid sequences can be accomplished using a mathematical algorithm. For example, a mathematical algorithm useful for comparing two sequences is the algorithm of Karlin and Altschul (1990, Proc. Natl. Acad. Sci. USA 87:2264-2268), modified as in Karlin and Altschul (1993, Proc. Natl. Acad. Sci. USA 90:5873-5877). This algorithm is incorporated into the NBLAST and XBLAST

programs of Altschul et al. (1990, J. Mol. Biol. 215:403-410), and can be accessed, for example, at the BLAST site of the National Center for Biotechnology Information (NCBI) world wide web site at the National Library of Medicine (NLM) at the National Institutes of Health (NIH). BLAST nucleotide searches can be performed with the NBLAST program (designated "blastn" at the NCBI web site), using the following parameters: gap penalty = 5; gap extension penalty = 2; mismatch penalty = 3; match reward = 1; expectation value 10.0; and word size = 11 to obtain nucleotide sequences homologous to a nucleic acid described herein. BLAST protein searches can be performed with the XBLAST program (designated "blastn" at the NCBI web site) or the NCBI "blastp" program, using the following parameters: expectation value 10.0, BLOSUM62 scoring matrix to obtain amino acid sequences homologous to a protein molecule described herein.

[0092] To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997, Nucleic Acids Res. 25:3389-3402). Alternatively, PSI-Blast or PHI-Blast can be used to perform an iterated search which detects distant relationships between molecules (*id.*) and relationships between molecules which share a common pattern. When utilizing BLAST, Gapped BLAST, PSI-Blast, and PHI-Blast programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used as available on the website of the National Center for Biotechnology Information of the National Library of Medicine at the National Institutes of Health.

[0093] The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, typically exact matches are counted.

[0094] "Polypeptide" refers to a polymer composed of amino acid residues, related naturally occurring structural variants, and synthetic non-naturally occurring analogs thereof linked via peptide bonds, related naturally occurring structural variants, and synthetic non-naturally occurring analogs thereof. Synthetic polypeptides can be synthesized, for example, using an automated polypeptide synthesizer. A "polypeptide," as the term is used herein, therefore refers to any size polymer of amino acid residues, provided that the polymer contains at least two amino acid residues.

[0095] The term "protein" typically refers to large peptides, also referred to herein as "polypeptides." The term "peptide" typically refers to short polypeptides. However, the terms "peptide," "protein" and "polypeptide" are used interchangeably herein. For example,

the term "peptide" may refer to an amino acid polymer of three amino acids, as well as an amino acid polymer of several hundred amino acids.

[0096] As used herein, amino acids are represented by the full name thereof, by the three letter code corresponding thereto, or by the one-letter code corresponding thereto, as indicated in the following table:

Full Name	Three-Letter Code	One-Letter Code
Aspartic Acid	Asp	D
Glutamic Acid	Glu	E
Lysine	Lys	K
Arginine	Arg	R
Histidine	His	H
Tyrosine	Tyr	Y
Cysteine	Cys	C
Asparagine	Asn	N
Glutamine	Gln	Q
Serine	Ser	S
Threonine	Thr	T
Glycine	Gly	G
Alanine	Ala	A
Valine	Val	V
Leucine	Leu	L
Isoleucine	Ile	I
Methionine	Met	M
Proline	Pro	P
Phenylalanine	Phe	F
Tryptophan	Trp	W

[0097] Conventional notation is used herein to portray polypeptide sequences: the left-hand end of a polypeptide sequence is the amino-terminus; the right-hand end of a polypeptide sequence is the carboxyl-terminus.

[0098] A "therapeutic peptide" as the term is used herein refers to any peptide that is useful to treat a disease state or to improve the overall health of a living organism. A therapeutic

peptide may effect such changes in a living organism when administered alone, or when used to improve the therapeutic capacity of another substance. The term "therapeutic peptide" is used interchangeably herein with the terms "therapeutic polypeptide" and "therapeutic protein."

[0099] A "reagent peptide" as the term is used herein refers to any peptide that is useful in food biochemistry, bioremediation, production of small molecule therapeutics, and even in the production of therapeutic peptides. Typically, reagent peptides are enzymes capable of catalyzing a reaction to produce a product useful in any of the aforementioned areas. The term "reagent peptide" is used interchangeably herein with the terms "reagent polypeptide" and "reagent protein."

[0100] A "glycopeptide" as the term is used herein refers to a peptide having at least one carbohydrate moiety covalently linked thereto. It will be understood that a glycopeptide may be a "therapeutic glycopeptide," as described above. The term "glycopeptide" is used interchangeably herein with the terms "glycopolypeptide" and "glycoprotein."

[0101] A "vector" is a composition of matter which comprises an isolated nucleic acid and which can be used to deliver the isolated nucleic acid to the interior of a cell. Numerous vectors are known in the art including, but not limited to, linear nucleic acids, nucleic acids associated with ionic or amphiphilic compounds, plasmids, and viruses. Thus, the term "vector" includes an autonomously replicating plasmid or a virus. The term should also be construed to include non-plasmid and non-viral compounds which facilitate transfer of nucleic acid into cells, such as, for example, polylysine compounds, liposomes, and the like. Examples of viral vectors include, but are not limited to, adenoviral vectors, adeno-associated virus vectors, retroviral vectors, and the like.

[0102] "Expression vector" refers to a vector comprising a recombinant nucleic acid comprising expression control sequences operatively linked to a nucleotide sequence to be expressed. An expression vector comprises sufficient cis-acting elements for expression; other elements for expression can be supplied by the host cell or in an in vitro expression system. Expression vectors include all those known in the art, such as cosmids, plasmids (e.g., naked or contained in liposomes) and viruses that incorporate the recombinant nucleic acid.

[0103] A "multiple cloning site" as the term is used herein is a region of a nucleic acid vector that contains more than one sequence of nucleotides that is recognized by at least one restriction enzyme.

[0104] An "antibiotic resistance marker" as the term is used herein refers to a sequence of nucleotides that encodes a protein which, when expressed in a living cell, confers to that cell the ability to live and grow in the presence of an antibiotic.

[0105] As used herein, the term "GalNAcT2" refers to N-acetyl-D-galactosamine transferase 2.

[0106] As the term is used herein, a "truncated" form of a peptide refers to a peptide that is lacking one or more amino acid residues as compared to the full-length amino acid sequence of the peptide. For example, the peptide "NH2-Ala-Glu-Lys-Leu-COOH" is an N-terminally truncated form of the full-length peptide "NH2-Gly-Ala-Glu-Lys-Leu-COOH." The terms "truncated form" and "truncation mutant" are used interchangeably herein. By way of a non-limiting example, a truncated peptide is a GalNAcT2 polypeptide comprising an active domain, a stem domain, a transmembrane domain, and a signal domain, wherein the signal domain is lacking a single N-terminal amino acid residue as compared to the full length GalNAcT2.

[0107] The term "saccharide" refers in general to any carbohydrate, a chemical entity with the most basic structure of $(CH_2O)_n$. Saccharides vary in complexity, and may also include nucleic acid, amino acid, or virtually any other chemical moiety existing in biological systems.

[0108] "Monosaccharide" refers to a single unit of carbohydrate of a defined identity.

[0109] "Oligosaccharide" refers to a molecule consisting of several units of carbohydrates of defined identity. Typically, saccharide sequences between 2-20 units may be referred to as oligosaccharides.

[0110] "Polysaccharide" refers to a molecule consisting of many units of carbohydrates of defined identity. However, any saccharide of two or more units may correctly be considered a polysaccharide.

[0111] As used herein, a saccharide "donor" is a moiety that can provide a saccharide to a glycosyltransferase so that the glycosyltransferase may transfer the saccharide to a saccharide acceptor. By way of a non-limiting example, a GalNAc donor may be UDP-GalNAc.

[0112] As used herein, a saccharide "acceptor" is a moiety that can accept a saccharide from a saccharide donor. A glycosyltransferase can covalently couple a saccharide to a saccharide acceptor. By way of a non-limiting example, G-CSF may be a GalNAc acceptor, and a GalNAc moiety may be covalently coupled to a GalNAc acceptor by way of a GalNAc-transferase. In some embodiments, a saccharide acceptor is a protein or peptide comprising an O glycosylation site. In further embodiments, saccharide acceptors include, *e.g.*, erythropoietin, human growth hormone, granulocyte colony stimulating factor, interferons alpha, -beta, and –gamma, Factor IX, follicle stimulating hormone, interleukin-2, erythropoietin, anti-TNF-alpha, and a lysosomal hydrolase

[0113] An oligosaccharide with a "defined size" is one which consists of an identifiable number of monosaccharide units. For example, an oligosaccharide consisting of 10 monosaccharide units is one which may consist of 10 identical monosaccharide units or 5 monosaccharide units of a first identity and 5 monosaccharide units of a second identity. Further, an oligosaccharide of defined size that consists of monosaccharide units of heterogeneous identity may have the monosaccharide units in any order from beginning to end of the oligosaccharide.

[0114] An oligosaccharide of "random size" is one which may be synthesized using methods that do not provide oligosaccharide products of defined size. For example, a method of oligosaccharide synthesis may provide oligosaccharides that range from two monosaccharide units to twenty-two saccharide units, including any or all lengths in between.

[0115] "Commercial scale" refers to gram scale production of a product saccharide, or glycoprotein, or glycopeptide in a single reaction. In preferred embodiments, commercial scale refers to production of greater than about 50, 75, 80, 90 or 100, 125, 150, 175, or 200 grams.

[0116] The term "sialic acid" refers to any member of a family of nine-carbon carboxylated sugars. The most common member of the sialic acid family is N-acetyl-neuraminic acid (2-keto-5-acetamido-3,5-dideoxy-D-glycero-D-galactononulopyranos-1-onic acid (often abbreviated as Neu5Ac, NeuAc, or NANA). A second member of the family is N-glycolyl-neuraminic acid (Neu5Gc or NeuGc), in which the N-acetyl group of NeuAc is hydroxylated.

A third sialic acid family member is 2-keto-3-deoxy-nonulosonic acid (KDN) (Nadano *et al.* (1986) *J. Biol. Chem.* **261**: 11550-11557; Kanamori *et al.*, *J. Biol. Chem.* **265**: 21811-21819 (1990)). Also included are 9-substituted sialic acids such as a 9-O-C_1-C_6 acyl-Neu5Ac like 9-O-lactyl-Neu5Ac or 9-O-acetyl-Neu5Ac, 9-deoxy-9-fluoro-Neu5Ac and 9-azido-9-deoxy-Neu5Ac. For review of the sialic acid family, *see*, *e.g.*, Varki, *Glycobiology* **2**: 25-40 (1992); *Sialic Acids: Chemistry, Metabolism and Function*, R. Schauer, Ed. (Springer-Verlag, New York (1992)). The synthesis and use of sialic acid compounds in a sialylation procedure is disclosed in international application WO 92/16640, published October 1, 1992.

[0117] A " method of remodeling a protein, a peptide, a glycoprotein, or a glycopeptide" as used herein, refers to addition of a sugar residue to a protein, a peptide, a glycoprotein, or a glycopeptide using a glycosyltransferase. In a preferred embodiment, the sugar residue is covalently attached to a PEG molecule.

[0118] An "unpaired cysteine residue" as used herein, refers to a cysteine residue, which in a correctly folded protein (*i.e.*, a protein with biological activity), does not form a disulfide bind with another cysteine residue.

[0119] An "insoluble glycosyltransferase" refers to a glycosyltransferase that is expressed in bacterial inclusion bodies. Insoluble glycosyltransferases are typically solubilized or denatured using *e.g.*, detergents or chaotropic agents or some combination. "Refolding" refers to a process of restoring the structure of a biologically active glycosyltransferase to a glycosyltransferase that has been solubilized or denatured. Thus, a refolding buffer, refers to a buffer that enhances or accelerates refolding of a glycosyltransferase.

[0120] A "redox couple" refers to mixtures of reduced and oxidized thiol reagents and include reduced and oxidized glutathione (GSH/GSSG), cysteine/cystine, cysteamine/cystamine, DTT/GSSG, and DTE/GSSG. (*See, e.g.*, Clark, *Cur. Op. Biotech.* 12:202-207 (2001)).

[0121] The term “contacting” is used herein interchangeably with the following: combined with, added to, mixed with, passed over, incubated with, flowed over, etc.

[0122] The term "PEG" refers to poly(ethylene glycol). PEG is an exemplary polymer that has been conjugated to peptides. The use of PEG to derivatize peptide therapeutics has been demonstrated to reduce the immunogenicity of the peptides and prolong the clearance time from the circulation. For example, U.S. Pat. No. 4,179,337 (Davis *et al.*) concerns non-

immunogenic peptides, such as enzymes and peptide hormones coupled to polyethylene glycol (PEG) or polypropylene glycol. Between 10 and 100 moles of polymer are used per mole peptide and at least 15% of the physiological activity is maintained.

[0123] The term "specific activity" as used herein refers to the catalytic activity of an enzyme, *e.g.*, a recombinant glycosyltransferase fusion protein of the present invention, and may be expressed in activity units. As used herein, one activity unit catalyzes the formation of 1 μmol of product per minute at a given temperature (*e.g.*, at 37°C) and pH value (*e.g.*, at pH 7.5). Thus, 10 units of an enzyme is a catalytic amount of that enzyme where 10 μmol of substrate are converted to 10 μmol of product in one minute at a temperature of, *e.g.*, 37 °C and a pH value of, *e.g.*, 7.5.

[0124] "N-linked" oligosaccharides are those oligosaccharides that are linked to a peptide backbone through asparagine, by way of an asparagine-N-acetylglucosamine linkage. N-linked oligosaccharides are also called "N-glycans." All N-linked oligosaccharides have a common pentasaccharide core of $Man_3GlcNAc_2$. They differ in the presence of, and in the number of branches (also called antennae) of peripheral sugars such as N-acetylglucosamine, galactose, N-acetylgalactosamine, fucose and sialic acid. Optionally, this structure may also contain a core fucose molecule and/or a xylose molecule.

[0125] "O-linked" oligosaccharides are those oligosaccharides that are linked to a peptide backbone through threonine, serine, hydroxyproline, tyrosine, or other hydroxy-containing amino acids.

[0126] The term "substantially" in the above definitions of "substantially uniform" generally means at least about 60%, at least about 70%, at least about 80%, or more preferably at least about 90%, and still more preferably at least about 95% of the acceptor substrates for a particular glycosyltransferase are glycosylated.

[0127] A "fusion protein" refers to a protein comprising amino acid sequences that are in addition to, in place of, less than, and/or different from the amino acid sequences encoding the original or native full-length protein or subsequences thereof.

[0128] A "stem region" with reference to glycosyltransferases refers to a protein domain, or a subsequence thereof, which in the native glycosyltransferases is located adjacent to the trans-membrane domain, and has been reported to function as a retention signal to maintain the glycosyltransferase in the Golgi apparatus and as a site of proteolytic cleavage. Stem

regions generally start with the first hydrophilic amino acid following the hydrophobic transmembrane domain and end at the catalytic domain, or in some cases the first cysteine residue following the transmembrane domain. Exemplary stem regions include, but is not limited to, the stem region of eukaryotic ST6GalNAcI, amino acid residues from about 30 to about 207 (see *e.g.*, the murine enzyme), amino acids 35-278 for the h uman enzyme or amino acids 37-253 for the chicken enzyme; the stem region of mammalian GalNAcT2, amino acid residues from about 71 to about 129 (see *e.g.*, the rat enzyme).

[0129] A "catalytic domain" refers to a protein domain, or a subsequence thereof, that catalyzes an enzymatic reaction performed by the enzyme. For example, a catalytic domain of a sialyltransferase will include a subsequence of the sialyltransferase sufficient to transfer a sialic acid residue from a donor to an acceptor saccharide. A catalytic domain can include an entire enzyme, a subsequence thereof, or can include additional amino acid sequences that are not attached to the enzyme, or a subsequence thereof, as found in nature.

[0130] The term "isolated" refers to material that is substantially or essentially free from components which interfere with the activity of an enzyme. For a saccharide, protein, or nucleic acid of the invention, the term "isolated" refers to material that is substantially or essentially free from components which normally accompany the material as found in its native state. Typically, an isolated saccharide, protein, or nucleic acid of the invention is at least about 80% pure, usually at least about 90%, and preferably at least about 95% pure as measured by band intensity on a silver stained gel or other method for determining purity. Purity or homogeneity can be indicated by a number of means well known in the art. For example, a protein or nucleic acid in a sample can be resolved by polyacrylamide gel electrophoresis, and then the protein or nucleic acid can be visualized by staining. For certain purposes high resolution of the protein or nucleic acid may be desirable and HPLC or a similar means for purification, for example, may be utilized.

Description

I. Isolated nucleic acids

A. Generally

[0131] Exemplified herein are various truncation mutants of human GalNAcT2. However, the present invention should not be construed to cover a human GalNAcT2 truncation mutant polypeptide lacking amino acid residues 1-51.

[0132] Full-length GalNAcT2 nucleic acids encode polypeptides that have a domain structure similar to other glycosyltransferases, including an N-terminal signal domain, a transmembrane domain, a stem domain, and an active domain, wherein the active domain may comprise the majority of the amino acid sequence of such polypeptides. As will be understood by one of skill in the art, the presence of domain structure(s) extraneous to the active domain of recombinant GalNAcT2 polypeptides may have a negative effect on the solubility, stability and activity of the polypeptide in an aqueous or in vitro environment. For example, while not wishing to be bound by any particular theory, the presence of a hydrophobic transmembrane domain on a recombinant GalNAcT2 polypeptide used in an in vitro reaction mixture may render the polypeptide less soluble than a recombinant GalNAcT2 polypeptide without a hydryophobic transmembrane domain, and further, may even decrease the enzymatic activity of the polypeptide by affecting or destabilizing the folded structure.

[0133] Therefore, it is desirable to produce recombinant GalNAcT2 nucleic acids that encode GalNAcT2 that is shorter than full-length GalNAcT2, for the purpose of enhancing the activity, stability and/or utility of GalNAcT2 polypeptides. The present invention provides such modified forms of GalNAcT2. More particularly, the present invention provides isolated nucleic acids encoding such truncated polypeptides.

[0134] Nucleic acids of the present invention encode truncated forms of GalNacT2 polypeptides, as described in greater detail elsewhere herein. A truncated GalNAcT2 polypeptide encoded by a nucleic acid of the present invention, also referred to herein as a "truncation mutant," may be truncated in various ways, as would be understood by the skilled artisan. Examples of truncated polypeptides encoded by a nucleic acid of the present invention include, but are not limited to, a polypeptide lacking a single N-terminal residue, a polypeptide lacking a single C-terminal residue, a polypeptide lacking both an single N-terminal residue and a single C-terminal residue, a polypeptide lacking a contiguous sequence of residues from the N-terminus, a polypeptide lacking a contiguous sequence of residues from the C-terminus, and any combinations thereof.

[0135] Therefore, it will be understood, based on the disclsure set forth herein, that truncations of nucleic acids encoding GalNAcT2 polypeptides may be made for numerous reasons. In one embodiment of the invention, a truncation may be made in order to remove part or all of the nucleic acid sequence encoding the signal peptide domain of an GalNAcT2.

[0136] In another embodiment of the invention, a truncation may be made in order to remove part or all of a nucleic acid sequence encoding a transmembrane domain of an GalNAcT2. By way of a non-limiting example, removal of a part or all of a nucleic acid sequence encoding a transmembrane domain may increase the solubility or stability of the encoded GalNAcT2 polypeptide and/or may increase the level of expression of the encoded polypeptide.

[0137] In yet another embodiment of the invention, a truncation may be made in order to remove part or all of a nucleic acid sequence encoding a stem domain of an GalNAcT2. By way of a non-limiting example, removal of a part or all of a nucleic acid sequence encoding a stem domain may increase the solubility or stability of the encoded GalNAcT2 polypeptide and/or may increase the level of expression of the encoded polypeptide.

[0138] The skilled artisan, when armed with the disclosure set forth herein, will understand how to design and create a truncation mutant of GalNAcT2 as set forth in detail elsewhere herein. In one aspect of the invention, the nucleic acid residue at which a truncation is made may be a highly-conserved residue. In another aspect of the invention, the nucleic acid residue at which a truncation is made may be selected such that the encoded polypeptide has a new N-terminal amino acid residue that will aid in the purification of the expressed polypeptide. In yet another aspect, the nucleic acid residue at which a truncation is made may be selected such that the encoded truncated polypeptide does not contain a specific secondary and/or tertiary structure.

B. GalNAcT2 Isolated Nucleic Acids

[0139] The present invention features nucleic acids encoding smaller than full-length GalNAcT2. That is, the present invention features a nucleic acid encoding a truncated GalNAcT2 polypeptide, provided the polypeptide expressed by the nucleic acid retains the biological activity of the full-length protein. In one aspect of the invention, a truncated polypeptide is a human truncated GalNAcT2 polypeptide.

[0140] As would be understood by the skilled artisan, a nucleic acid encoding a full-length human GalNAcT2 may contain a nucleic acid sequence encoding one or more identifyable polypeptide domains in addition to the "active domain," the domain primarily responsible for the catalytic activity, of GalNAcT2. This is because it is known in that art that a full-length GalNAcT2 polypeptide, and in particular, a full-length human GalNAcT2 polypeptide, contains a signal domain, a transmembrane domain, and a stem domain, in addition to an

active domain. Accordingly, a nucleic acid encoding a full-length human GalNAcT2 may encode a polypeptide that has a signal domain at the amino-terminus of the polypeptide, followed by a transmembrane domain immediately adjacent to the signal domain, followed by a stem domain that is immediately adjacent to the transmembrane domain, followed by an active domain that extends to the carboxy-terminus of the polypeptide and is located immediately adjacent to the stem domain.

[0141] Therefore, in one embodiment, an isolated nucleic acid of the invention may encode a truncated human GalNAcT2 polypeptide, wherein the truncated human GalNAcT2 polypeptide is lacking all or a portion of the GalNAcT2 signal domain. In another embodiment, an isolated nucleic acid of the invention may encode a truncated human GalNAcT2 polypeptide, wherein the truncated human GalNAcT2 polypeptide is lacking the GalNAcT2 signal domain and all or a portion of the GalNAcT2 transmembrane domain. In yet another embodiment, a nucleic acid of the invention may encode a truncated human GalNAcT2 polypeptide, wherein the truncated human GalNAcT2 polypeptide is lacking the GalNAcT2 signal domain, the GalNAcT2 transmembrane domain and all or a portion the GalNAcT2 stem domain.

[0142] When armed with the disclosure of the present invention, the skilled artisan will know how to make and use these and other such truncation mutants of human GalNAcT2.

[0143] The "biological activity of GalNAcT2" is the ability to transfer a GalNAc moiety from a GalNAc donor to an acceptor molecule. Full-length human GalNAcT2, the sequence of which is set forth in SEQ ID NO:1, exhibits such activity. The "biological activity of a GalNAcT2 truncated polypeptide" is similarly the ability to transfer a GalNAc moiety from a GalNAc donor to an acceptor molecule. That is, a truncated GalNAcT2 polypeptide of the present invention can catalyze the same glycosyltransfer reaction as the full-length GalNAcT2. By way of a non-limiting example, a truncated human GalNAcT2 polypeptide encoded by a GalNAcT2 nucleic acid of the invention has the ability to transfer a GalNAc moiety from a UDP-GalNAc donor to a granulocyte-colony stimulating factor (G-CSF) acceptor, wherein such a transfer results in the O-linked covalent coupling of a GalNAc moiety to a threonine residue of G-CSF.

[0144] Therefore, a nucleic acid encoding a smaller than full-length, or "truncated," GalNAcT2 is included in the present invention provided that the truncated GalNAcT2 has GalNAcT2 biological activity.

[0145] The methods and compositions of the invention should not be construed to be limited solely to a nucleic acid comprising a GalNAcT2 truncation mutant as disclosed herein, but rather, should be construed to encompass any nucleic acid encoding a GalNAcT2 truncated mutant, prepared in accordance with the disclosure herein, either known or unknown, which is capable of catalyzing transfer of a GalNAc to a GalNAc acceptor. Modified nucleic acid sequences, i.e. nucleic acid sequences having sequences that differ from the nucleic acid sequences encoding the naturally-occurring proteins, are also encompassed by methods and compositions of the invention, so long as the modified nucleic acid still encodes a truncated protein having the biological activity of catalyzing the transfer of a GalNAc to a GalNAc acceptor, for example. These modified nucleic acid sequences include modifications caused by point mutations, modifications due to the degeneracy of the genetic code or naturally occurring allelic variants, and further modifications that have been introduced by genetic engineering, i.e., by the hand of man. Thus, the term nucleic acid also specifically includes nucleic acids composed of bases other than the five biologically occurring bases (adenine, guanine, thymine, cytosine and uracil).

[0146] The present invention features an isolated nucleic acid comprising a nucleic acid sequence that is at least about 90%, 95%, 97%, 98%, or 99% identical to a nucleic acid sequence set forth in any one of SEQ ID NO:3, SEQ ID NO:7 or SEQ ID NO:9. The present invention also features an isolated nucleic acid sequence comprising any one of the sequences set forth in SEQ ID NO:3, SEQ ID NO:7 or SEQ ID NO:9, wherein the isolated nucleic acid encodes a truncated GalNAcT2 polypeptide.

[0147] The present invention also encompasses isolated nucleic acid molecules encoding a truncated GalNAcT2 polypeptide that contains changes in amino acid residues that are not essential for activity. Such polypeptides encoded by an isolated nucleic acid of the invention differ in amino acid sequence from any one of the sequences set forth in SEQ ID NO:4, SEQ ID NO:8 or SEQ ID NO:10, yet retain the biological activity of GalNAcT2. By way of a non-limiting example, an isolated nucleic acid of the invention may include a nucleotide sequence encoding a polypeptide having an amino acid sequence that is at least about 90%, 95%, 97%, 98%, or 99% identical to the amino acid sequence of SEQ ID NO:4. Further, by way of another non-limiting example, an isolated nucleic acid of the invention may include a nucleotide sequence encoding a polypeptide that has an amino acid sequence at least about 90%, 95%, 97%, 98%, or 99% identical to an amino acid sequence set forth in any one of SEQ ID NO:8 or SEQ ID NO:10.

[0148] The determination of percent identity between two nucleotide or amino acid sequences can be accomplished using a mathematical algorithm. For example, a mathematical algorithm useful for comparing two sequences is the algorithm of Karlin and Altschul (1990, Proc. Natl. Acad. Sci. USA 87:2264-2268), modified as in Karlin and Altschul (1993, Proc. Natl. Acad. Sci. USA 90:5873-5877). This algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990, J. Mol. Biol. 215:403-410), and can be accessed, for example at the National Center for Biotechnology Information (NCBI) world wide web site. BLAST nucleotide searches can be performed with the NBLAST program (designated "blastn" at the NCBI web site), using the following parameters: gap penalty = 5; gap extension penalty = 2; mismatch penalty = 3; match reward = 1; expectation value 10.0; and word size = 11 to obtain nucleotide sequences homologous to a nucleic acid described herein. BLAST protein searches can be performed with the XBLAST program (designated "blastn" at the NCBI web site) or the NCBI "blastp" program, using the following parameters: expectation value 10.0, BLOSUM62 scoring matrix to obtain amino acid sequences homologous to a protein molecule described herein. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997, Nucleic Acids Res. 25:3389-3402). Alternatively, PSI-Blast or PHI-Blast can be used to perform an iterated search which detects distant relationships between molecules and relationships between molecules which share a common pattern. When utilizing BLAST, Gapped BLAST, PSI-Blast, and PHI-Blast programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See, generally, the internet website for the National Center for Biotechnology Information, which is maintained by the National Library of Medicine and the National Institutes of Health.

[0149] In another aspect, a nucleic acid useful in the methods and compositions of the present invention and encoding a truncated GalNAcT2 polypeptide may have at least one nucleotide inserted into the nucleic acid sequence of such a truncated mutant. Alternatively, an additional nucleic acid encoding a truncated GalNAcT2 polypeptide may have at least one nucleotide deleted from the nucleic acid sequence. Further, a GalNAcT2 nucleic acid encoding a truncated mutant and useful in the invention may have both a nucleotide insertion and a nucleotide deletion present in a single nucleic acid sequence encoding the truncated polypeptide.

[0150] Techniques for introducing changes in nucleotide sequences that are designed to alter the functional properties of the encoded proteins or polypeptides are well known in the

art. Such modifications include the deletion, insertion, or substitution of bases, and thus, changes in the amino acid sequence. As is known to one of skill in the art, nucleic acid insertions and/or deletions may be designed into the gene for numerous reasons, including, but not limited to modification of nucleic acid stability, modification of nucleic acid expression levels, modification of expressed polypeptide stability or half-life, modification of expressed polypeptide activity, modification of expressed polypeptide properties and characteristics, and changes in glycosylation pattern. All such modifications to the nucleotide sequences encoding such proteins are encompassed by the present invention.

[0151] It is not intended that methods and compositions of the present invention be limited by the nature of the nucleic acid employed. The target nucleic acid encompassed by methods and compositions of the invention may be native or synthesized nucleic acid. The nucleic acid may be DNA or RNA and may exist in a double-stranded, single-stranded or partially double-stranded form. Furthermore, the nucleic acid may be found as part of a virus or other macromolecule. See, e.g., Fasbender et al., 1996, J. Biol. Chem. 272:6479-89.

II. Vectors and Expression Systems

[0152] In other related aspects, the invention includes an isolated nucleic acid encoding a truncated GalNAcT2 polypeptide operably linked to a nucleic acid comprising a promoter/regulatory sequence such that the nucleic acid is preferably capable of directing expression of the polypeptide encoded by the nucleic acid. Thus, the invention encompasses expression vectors and methods for the introduction of exogenous DNA into cells with concomitant expression of the exogenous DNA in those cells, as described, for example, in Sambrook et al. (Third Edition, 2001, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York), and in Ausubel et al. (1997, Current Protocols in Molecular Biology, John Wiley & Sons, New York).

[0153] Expression of a truncated GalNAcT2 polypeptide in a cell may be accomplished by generating a plasmid, viral, or other type of vector comprising a nucleic acid encoding the appropriate nucleic acid, wherein the nucleic acid is operably linked to a promoter/regulatory sequence which serves to drive expression of the encoded polypeptide, with or without tag, in cells in which the vector is introduced. In addition, promoters which are well known in the art which are induced in response to inducing agents such as metals, glucocorticoids, and the like, are also contemplated in the invention. Thus, it will be appreciated that the invention includes the use of any promoter/regulatory sequence, which is either known or unknown,

and which is capable of driving expression of the truncated GalNAcT2 polypeptide operably linked thereto.

[0154] In an expression system useful in the present invention, a nucleic acid encoding a truncated GalNAcT2 polypeptide may be fused to one or more additional nucleic acids encoding a functional polypeptide. By way of a non-limiting example, an affinity tag coding sequence may be inserted into a nucleic acid vector adjacent to, upstream from, or downstream from a truncated GalNAcT2 polypeptide coding sequence. As will be understood by one of skill in the art, an affinity tag will typically be inserted into a multiple cloning site in frame with the truncated GalNAcT2 polypeptide. One of skill in the art will also understand that an affinity tag coding sequence can be used to produce a recombinant fusion protein by concomitantly expressing the affinity tag and truncated GalNAcT2 polypeptide. The expressed fusion protein can then be isolated, purified, or identified by means of the affinity tag.

[0155] Affinity tags useful in the present invention include, but are not limited to, a maltose binding protein, a histidine tag, a Factor IX tag, a glutathione-S-transferase tag, a FLAG-tag, and a starch binding domain tag. Other tags are well known in the art, and the use of such tags in the present invention would be readily understood by the skilled artisan.

[0156] As would be understood by one of skill in the art, a vector comprising a truncated GalNAcT2 polypeptide of the present invention may be used to express the truncated polypeptide as either a non-fusion or as a fusion protein. Selection of any particular plasmid vector or other DNA vector is not a limiting factor in this invention and a wide plethora of vectors are well-known in the art. Further, it is well within the skill of the artisan to choose particular promoter/regulatory sequences and operably link those promoter/regulatory sequences to a DNA sequence encoding a truncated GalNAcT2 polypeptide. Such technology is well known in the art and is described, for example, in Sambrook et al. (Third Edition, 2001, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York), and in Ausubel et al. (1997, Current Protocols in Molecular Biology, John Wiley & Sons, New York). By way of a non-limiting example, a vector useful in one embodiment of the present invention is based on the pcWori+ vector (Muchmore et al., 1987, Meth. Enzymol. 177:44-73).

[0157] The invention thus includes a vector comprising an isolated nucleic acid encoding a truncated GalNAcT2 polypeptide. The incorporation of a nucleic acid into a vector and the

choice of vectors is well-known in the art as described in, for example, Sambrook et al. (Third Edition, 2001, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York), and in Ausubel et al. (1997, Current Protocols in Molecular Biology, John Wiley & Sons, New York).

[0158] In an aspect of the invention, an isolated nucleic acid encoding a truncated GalNAcT2 polypeptide is integrated into the genome of a host cell in conjunction with a nucleic acid encoding a truncated GalNAcT2 polypeptide. In another aspect of the invention, a cell is transiently transfected with an isolated nucleic acid encoding a truncated GalNAcT2 polypeptide. In yet another aspect of the invention, a cell is stably transfected with an isolated nucleic acid encoding a truncated GalNAcT2 polypeptide.

[0159] For the purpose of inserting an isolated nucleic acid into a cell, one of skill in the art would also understand that the methods available and the methods required to introduce an isolated nucleic acid of the invention into a host cell vary and depend upon the choice of host cell. Suitable methods of introducing an isolated nucleic acid into a host cell are well-known in the art. Other suitable methods for transforming or transfecting host cells may include, but are not limited to, those found in Sambrook, et al. (Molecular Cloning: A Laboratory Manual. 3rd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001), and other such laboratory manuals.

[0160] A nucleic acid encoding a truncated GalNAcT2 polypeptide may be purified by any suitable means, as are well known in the art. For example, the nucleic acids can be purified by reverse phase or ion exchange HPLC, size exclusion chromatography or gel electrophoresis. Of course, the skilled artisan will recognize that the method of purification will depend in part on the size of the DNA to be purified.

[0161] The present invention also features a recombinant bacterial host cell comprising , *inter alia*, a nucleic acid vector as described elsewhere herein. In one aspect, the recombinant cell is transformed with a vector of the present invention. The transformed vector need not be integrated into the cell genome nor does it need to be expressed in the cell. However, the transformed vector will be capable of being expressed in the cell. In one aspect of the invention, E. coli is used for transformation of a vector of the present invention and expression of protein therefrom. In another aspect of the invention, a K-12 strain of E. coli is useful for expression of protein from a vector of the present invention. Strains of E. coli

useful in the present invention include, but are not limited to, JM83, JM101, JM103, JM109, W3110, chi1776, and JA221.

[0162] It will be understood that a host cell useful in the present invention will be capable of growth and culture on a small scale, medium scale, or a large scale. For example, a host cell of the invention is useful for testing the expression of a protein from a vector of the invention equally as much as it is useful for large scale production of a reagent or therapeutic protein product. Techniques useful in culturing host cells and expressing protein from a vector contained therein are well known in the art and will therefore not be listed herein.

[0163] A host cell useful in methods of the present invention, as described above, may be prepared according to various methods, as would be understood by the skilled artisan when armend with the disclosure set forth herein. In one aspect, a host cell of the present invention may be transformed with a vector of the present invention to produce a transformed host cell of the invention. Transformation, as known to the skilled artisan, includes the process of inserting a nucleic acid vector into a host cell, such that the host cell containing the nucleic acid vector remains viable. Such transformation of nucleic acid into a bacterial cell is useful for purposes including, but not limited to, creation of a stably-transformed host cell, making a biological deposit, propagating the vector-containing host cell, propagating the vector-containing host cell for the production and isolation of additional vector, expression of target protein encoded by vector, and the like.

[0164] Methods of transforming a cell with a vector are numerous and well-known in the art, and will therefore not be listed here. By way of a non-limiting example, a competent bacterial cell of the invention may be transformed by a vector of the invention using electroporation. Methods of making bacterial cells "competent" are well-known in the art, and typically involve preparation of the bacterial cells so that the cells take up exogenous DNA. Similarly, methods of electroporation are known in the art, and detailed descriptions of such methods may be found, for example, in Sambrook et al. (1989, *supra*). The transformation of a competent cell with vector DNA may be also accomplished using chemical-based methods. One example of a well-known chemical-based method of bacterial transformation is described by Inoue, et al. (1990, Gene 96:23-28). Other methods of transformation will be known to the skilled artisan.

[0165] A transformed host cell of the present invention may be used to express a truncated GalNAcT2 polypeptide of the present invention. In an embodiment of the invention, a

transformed host cell contains a vector of the invention, which contains therein a nucleic acid sequence encoding an truncated polypeptide of the invention. The truncated polypeptide is expressed using any expression method known in the art (for example, IPTG). The expressed truncated polypeptide may be contained within the host cell, or it may be secreted from the host cell into the growth medium.

[0166] Methods for isolating an expressed polypeptide are well-known in the art, and the skilled artisan will know how to determine the best method for isolation of an expressed polypeptide based on the characteristics of any given host cell expression system. By way of a non-limiting example, an expressed polypeptide that is secreted from a host cell may be isolated from the growth medium. Isolation of a polypeptide from a growth medium may include removal of bacterial cells and cellular debris. By way of another non-limiting example, an expressed polypeptide that is contained within a host cell may be isolated from the host cell. Isolation of such an "intracellular" expressed polypeptide may include disruption of the host cell and removal of cellular debris from the resultant mixture. These methods are not intended to be exclusive representations of the present invention, but rather, are merely for the purposes of illustration of various applications of the present invention.

[0167] Purification of a truncated polypeptide expressed in accordance with the present invention may be effected by any means known in the art. The skilled artisan will know how to determine the best method for the purification of a polypeptide expressed in accordance with the present invention. A purification method will be chosen by the skilled artisan based on factors such as, but not limited to, the expression host, the contents of the crude extract of the polypeptide, the size of the polypeptide, the properties of the polypeptide, the desired end product of the polypeptide purification process, and the subsequent use of the end product of the polypeptide purification process.

[0168] In an embodiment of the invention, isolation or purification of a truncated polypeptide expressed in accordance with the present invention may not be desired. In an aspect of the present invention, an expressed polypeptide may be stored or transported inside the bacterial host cell in which the polypeptide was expressed. In another aspect of the invention, an expressed polypeptide may be used in a crude lysate form, which is produced by lysis of a host cell in which the polypeptide was expressed. In yet another embodiment of the invention, an expressed polypeptide may be partially isolated or partially purified according to any of the methods set forth or described herein. The skilled artisan will know

when it is not desirable to isolate or purify a polypeptide of the invention, and will be familiar with the techniques available for the use and preparation of such polypeptides.

[0169] When armed with the disclosure set forth herein, the skilled artisan would also know how to prepare a eukaryotic host cell of the invention. As set forth elsewhere herein, and as would be known to one of skill in the art based on the disclosure provided herein, an isolated nucleic acid encoding a truncated GalNAcT2 polypeptide may be introduced into a eukaryotic host cell, for example, using a lentivirus-based genomic integration or plasmid-based transfection (Sambrook et al., Third Edition, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York (2001)). In one embodiment of the invention, a eukaryotic host cell is a fungal cell. In another embodiment, a nucleic acid encoding a truncated polypeptide of the invention is cloned into a lentiviral vector containing a specific promoter sequence for expression of the truncated polypeptide. The truncated polypeptide-containing lentiviral vector is then used to transfect a host cell for expression of the truncated polypeptide. Methods of making and using lentiviral vectors, such as those useful in the present invention, are well-known in the art and are not described further herein.

[0170] In yet another embodiment, a nucleic acid encoding a truncated polypeptide of the invention is introduced into a host cell using a viral expression system. Viral expression systems are well-known in the art, and will not be described in detail herein. In one aspect of the invention, a viral expression system is a mammalian viral expression system. In another aspect of the invention, a viral expression system is a baculovirus expression system. Such viral expression systems are typically commercially available from numerous vendors.

[0171] The skilled artisan will know how to use a host cell-vector expression system for the expression of a truncated polypeptide of the invention. Appropriate cloning and expression vectors for use with eukaryotic hosts are described by Sambrook, et al., in Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor, N.Y. (2001), the disclosure of which is hereby incorporated in its entirety by reference.

[0172] Insect cells can also be used for expression of a truncated polypeptide of the present invention. In an aspect of the invention, Sf9, SF9^{+}, Sf21, High Five™ or Drosophila Schneider S2 cells can be used. In yet another aspect of the invention, a baculovirus, or a baculovirus/insect cell expression system can be used to express a truncated polypeptide of the invention using a pAcGP67, pFastBac, pMelBac, or pIZ vector and a polyhedrin, p10, or

OpIE3 actin promoter. In another aspect of the invention, a Drosophila expression system can be used with a pMT or pAC5 vector and an MT or Ac5 promoter.

[0173] A truncated GalNAcT2 polypeptide of the invention can also be expressed in mammalian cells. In an aspect of the invention, 294, HeLa, HEK, NSO, Chinese hamster ovary (CHO), Jurkat, or COS cells can be used to express a truncated polypeptide of the invention. In the case of a mammalian cell expression of a truncated polypeptide, a suitable vector such as pT-Rex, pSecTag2, pBudCE4.1, or pCDNA/His Max vector can be used, along with, for example, a CMV promoter. As will be understood by the skilled artisan, the choice of promoter, as well as methods and strategies for introducing one or more promoters into a host cell used for expressing a truncated GalNAcT2 polypeptide of the invention are well-known in the art, and will vary depending upon the host cell and expression system used.

[0174] Various mammalian cell culture systems can be employed to express recombinant protein. Non-limiting examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell tines. Mammalian expression vectors may comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

[0175] The methods available and the methods required to introduce any isolated nucleic acid of the invention into a host cell vary and depend upon the choice of the host cell, as would be understoody by one of skill in the art. Suitable methods of introducing an isolated nucleic acid into a host cell are well-known in the art. By way of a non-limiting example, vector DNA can be introduced into a eukaryotic cell using conventional transfection techniques. As used herein, the term "transfection" refers to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (Molecular Cloning:

A Laboratory Manual. 3nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001), and other such laboratory manuals.

[0176] For example, for stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these transformants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a truncated polypeptide of the invention or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

III. Polypeptides

[0177] A truncated GalNAcT2 polypeptide of the present invention may be truncated in various ways, as would be known and understood by the skilled artisan, when armed with the present disclosure. Examples of truncated polypeptides of the present invention include, but are not limited to, a polypeptide lacking a single N-terminal residue, a polypeptide lacking a single C-terminal residue, a polypeptide lacking both an single N-terminal residue and a single C-terminal residue, a polypeptide lacking a contiguous sequence of residues from the N-terminus, a polypeptide lacking a contiguous sequence of residues from the C-terminus, and any such combinations thereof.

[0178] As would be understood by the skilled artisan, a full-length human GalNAcT2 polypeptide may contain one or more identifyable polypeptide domains in addition to the "active domain," the domain primarily responsible for the catalytic activity, of GalNAcT2. This is because it is known in that art that a full-length GalNAcT2 polypeptide, and in particular, a full-length human GalNAcT2 polypeptide, contains a signal domain, a transmembrane domain, and a stem domain, in addition to an active domain. Accordingly, a full-length human GalNAcT2 may have a signal domain at the amino-terminus of the polypeptide, followed by a transmembrane domain immediately adjacent to the signal domain, followed by a stem domain that is immediately adjacent to the transmembrane

domain, followed by an active domain that extends to the carboxy-terminus of the polypeptide and is located immediately adjacent to the stem domain.

[0179] Therefore, in one embodiment, a GalNAcT2 polypeptide of the invention is a truncated human GalNAcT2 polypeptide lacking all or a portion of the GalNAcT2 signal domain. In another embodiment, a GalNAcT2 polypeptide of the invention is a truncated human GalNAcT2 polypeptide lacking the GalNAcT2 signal domain and all or a portion of the GalNAcT2 transmembrane domain. In yet another embodiment, a GalNAcT2 polypeptide of the invention is a truncated human GalNAcT2 polypeptide lacking the GalNAcT2 signal domain, the GalNAcT2 transmembrane domain and all or a portion the GalNAcT2 stem domain. When armed with the disclosure of the present invention, the skilled artisan will know how to make and use these and other such truncation mutants of human GalNAcT2.

[0180] The size and identity of a truncated GalNAcT2 mutant of the present invention is based on the point at which the full-length polypeptide is truncated. By way of a non-limiting example, a "Δ40 human truncated GalNAcT2" mutant of the invention refers to a truncated GalNAcT2 polypeptide of the invention in which amino acids 1 through 40, counting from the N-terminus of the full-length polypeptide, are deleted from the polypeptide. Therefore, the N-terminus of the Δ40 human truncated GalNAcT2 mutant begins with the amino acid residue that would be referred to as "amino acid 41" of the full-length polypeptide. This nomenclature applies to all truncated GalNAcT2 polypeptides of the invention, including human GalNAcT2.

[0181] The present invention therefore also includes an isolated polypeptide comprising a truncated GalNAcT2 polypeptide. Preferably, an isolated truncated GalNAcT2 polypeptide of the present invention has at least about 90% identity to a polypeptide having the amino acid sequence of any one of the sequences set forth in SEQ ID NO:4, SEQ ID NO:8 or SEQ ID NO:10. More preferably, the isolated polypeptide is about 95% identical, and even more preferably, about 98% identical, still more preferably, about 99% identical, and most preferably, the isolated polypeptide comprising a truncated GalNAcT2 polypeptide is identical to the polypeptide set forth in one of SEQ ID NO:4, SEQ ID NO:8 or SEQ ID NO:10.

[0182] The present invention also provides for analogs of polypeptides which comprise a truncated GalNAcT2 polypeptide as disclosed herein. Analogs can differ from naturally

occurring proteins or peptides by conservative amino acid sequence differences or by modifications which do not affect sequence, or by both.

[0183] For example, conservative amino acid changes may be made, which although they alter the primary sequence of the protein or peptide, do not normally alter its function. Conservative amino acid substitutions typically include substitutions within the following groups:

glycine, alanine;
valine, isoleucine, leucine;
aspartic acid, glutamic acid;
asparagine, glutamine;
serine, threonine;
lysine, arginine;
phenylalanine, tyrosine.

[0184] Modifications (which do not normally alter primary sequence) include in vivo, or in vitro chemical derivatization of polypeptides, e.g., acetylation, or carboxylation. Also included are modifications of glycosylation, e.g., those made by modifying the glycosylation patterns of a polypeptide during its synthesis and processing or in further processing steps; e.g., by exposing the polypeptide to enzymes which affect glycosylation, e.g., mammalian glycosylating or deglycosylating enzymes. Also embraced are sequences which have phosphorylated amino acid residues, e.g., phosphotyrosine, phosphoserine, or phosphothreonine.

[0185] Also included are polypeptides which have been modified using ordinary molecular biological techniques so as to improve their resistance to proteolytic degradation or to optimize solubility properties or to render them more suitable as a therapeutic agent. Analogs of such polypeptides include those containing residues other than naturally occurring L-amino acids, e.g., D-amino acids or non-naturally occurring synthetic amino acids. The peptides of the invention are not limited to products of any of the specific exemplary processes listed herein.

[0186] Fragments of a truncated GalNAcT2 polypeptide of the invention are included in the present invention, provided the fragment possesses the biological activity of the full-length polypeptide. That is, a truncated GalNAcT2 polypeptide of the present invention can

catalyze the same glycosyltransfer reaction as the full-length GalNAcT2. By way of a non-limiting example, a truncated human GalNAcT2 polypeptide has the ability to transfer a GalNAc moiety from a UDP-GalNAc donor to a granulocyte-colony stimulating factor (G-CSF) acceptor, wherein such a transfer results in the O-linked covalent coupling of a GalNAc moiety to a threonine residue of G-CSF. Therefore, a smaller than full-length, or "truncated," GalNAcT2 is included in the present invention provided that the truncated GalNAcT2 has GalNAcT2 biological activity.

[0187] In another aspect of the present invention, compositions comprising an isolated truncated GalNAcT2 polypeptide as described herein may include highly purified truncated GalNAcT2 polypeptides. Alternatively, compositions comprising truncated GalNAcT2 polypeptides may include cell lysates prepared from the cells used to express the particular truncated GalNAcT2 polypeptides. Further, truncated GalNAcT2 polypeptides of the present invention may be expressed in one of any number of cells suitable for expression of polypeptides, such cells being well-known to one of skill in the art, as described in detail elsewhere herein.

[0188] Substantially pure protein isolated and obtained as described herein may be purified by following known procedures for protein purification, wherein an immunological, enzymatic or other assay is used to monitor purification at each stage in the procedure. Protein purification methods are well known in the art, and are described, for example in Deutscher *et al.* (ed., 1990, Guide to Protein Purification, Harcourt Brace Jovanovich, San Diego).

IV. Methods

[0189] The present invention features a method of expressing a truncated polypeptide. Polypeptides which can be expressed according to the methods of the present invention include a truncated GalNAcT2 polypeptide. More preferably, polypeptides which can be expressed according to the methods of the present invention include, but are not limited to, a truncated human GalNAcT2 polypeptide. In a preferred embodiment, a polypeptide which can be expressed according to the methods of the present invention is a polypeptide comprising any one of the polypeptide sequences set forth in SEQ ID NO:4, SEQ ID NO:8 or SEQ ID NO:10.

[0190] In one embodiment, the present invention features a method of expressing a truncated GalNAcT2 polypeptide encoded by an isolated nucleic acid of the invention, as

described elsewhere herein, wherein the expressed truncated GalNAcT2 polypeptide has the property of catalyzing the transfer of a GalNAc moiety to an acceptor moiety. In one aspect of the invention, a method of expressing a truncated GalNAcT2 polypeptide includes the steps of cloning an isolated nucleic acid of the invention into an expression vector, inserting the expression vector construct into a host cell, and expressing a truncated GalNAcT2 polypeptide therefrom.

[0191] Methods of expression of polypeptides, as well as construction of expression systems and recombinant host cells for expression of polypeptides, are discussed in extensive detail elsewhere herein. Methods of expression of a truncated polypeptide of the present invention will be understood to include, but not to be limited to, all such methods as described herein. In some expression systems, the truncated GalNAcT2 polypeptides of the invention are expressed as insoluble proteins, *e.g.*, in an inclusion protein in a bacterial host cell. Methods of refolding insoluble glycosyltransferases, including GalNAcT2 polypeptides, are disclosed in U.S. Provisional Patent Application Serial No. 60/542,210, filed February 4, 2004; U.S. Provisional Patent Application Serial No. 60/599,406, filed August 6, 2004; U.S. Provisional Patent Application Serial No. 60/627,406, filed November 12, 2004; and International Patent Application No. PCT/US05/03856, filed February 4, 2005; each of which are herein incorporated by reference for all purposes.

[0192] The present invention also features a method of catalyzing the transfer of a GalNAc moiety to a GalNAc acceptor moiety, wherein the GalNAc-transfer reaction is carried out by incubating a truncated GalNAcT2 polypeptide of the invention with a GalNAc donor moiety and a GalNAc acceptor moiety. In one aspect, a truncated GalNAcT2 polypeptide of the invention mediates the covalent linkage of a GalNAc moiety to a GalNAc acceptor moiety, thereby catalyzing the transfer of a GalNAc moiety to an acceptor moiety.

[0193] In one embodiment of the invention, a truncated GalNAcT2 polypeptide useful in a glycosyltransfer reaction is a truncated human GalNAcT2 polypeptide. In one aspect, the human GalNAcT2 glycosyltransfer reaction involves the transfer of a GalNAc residue from a GalNAc donor to a GalNAc acceptor.

[0194] By way of a non-limiting example, a method of catalyzing the transfer of a GalNAc moiety to an acceptor moiety includes the steps of incubating a truncated GalNAcT2 polypeptide with UDP-GalNAc GalNAc donor and a granulocyte colony stimulating factor

(G-CSF) acceptor moiety, wherein the truncated GalNAcT2 polypeptide mediates the transfer of GalNAc from the UDP-GalNAc donor to the GCSF acceptor.

[0195] Therefore, in one embodiment, the present invention also features a polypeptide acceptor moiety. In one embodiment of the invention, a polypeptide acceptor moiety is a human growth hormone. In another embodiment, a polypeptide acceptor moiety is an erythropoietin. In yet another embodiment, a polypeptide acceptor moiety is an interferon-alpha. In another embodiment, a polypeptide acceptor moiety is an interferon-beta. In another embodiment of the invention, a polypeptide acceptor moiety is an interferon-gamma. In still another embodiment of the invention, a polypeptide acceptor moiety is a lysosomal hydrolase. In another embodiment, a polypeptide acceptor moiety is a blood factor polypeptide. In still another embodiment, a polypeptide acceptor moiety is an anti-tumor necrosis factor-alpha. In another embodiment of the invention, a polypeptide acceptor moiety is follicle stimulating hormone.

[0196] In one embodiment, the present invention also features a method of transferring a GalNAc-polyethyleneglycol conjugate to an acceptor molecule. In one aspect, an acceptor molecule is a polypeptide. In another aspect, an acceptor molecule is a glycopeptide. Compositions and methods useful for designing, producing and transferring a GalNAc-polyethyleneglycol conjugate to an acceptor molecule are discussed at length in International (PCT) Patent Application No. WO03/031464 (PCT/US02/32263) and U.S. Patent Application No. 2004/0063911, each of which is incorporated herein by reference in its entirety. Methods of assaying for glycosyltransferase activity are well-known in the art. Various assays for detecting glycosyltransferases which can be used in accordance with the invention have been published. The following are illustrative, but should not be considered limiting, of those assays useful for detecting glycosyltransferase activity. Furukawa *et al.* (1985, *Biochem. J.*, 227:573-582) describe a borate-impregnated paper electrophoresis assay and a fluorescence assay. Roth *et al.* (1983, *Exp'l Cell Research* 143:217-225) describe application of the borate assay to glucuronyl transferases, previously assayed calorimetrically. Benau *et al.* (1990, *J. Histochem. Cytochem.*, 38:23-30) describe a histochemical assay based on the reduction, by NADH, of diazonium salts. See also U.S. Patent No. 6,284,493 of Roth, incorporated herein by reference.

EXPERIMENTAL EXAMPLES

[0197] The invention is now described with reference to the following examples. These examples are provided for the purpose of illustration only and the invention should in no way be construed as being limited to these examples but rather should be construed to encompass any and all variations which become evident as a result of the teaching provided herein.

Example 1: Cloning, Expression, and Refolding of Human Polypeptide N-acetylgalactosaminyltransferase II (GalNAcT2) in *E. coli* JM109

[0198] Four constructs were designed and created in order to assess the sialyltransferase activity of truncation mutants of human GalNAcT2. The four mutants created included Δ40, a truncation mutant which has as its new N-terminal residue an lysine that corresponds to R41 of the full-length human GalNAcT2 set forth in SEQ ID NO:2, Δ51, a truncation mutant which has as its new N-terminal residue an lysine that corresponds to K52 of the full-length human GalNAcT2 set forth in SEQ ID NO:2, Δ73, a truncation mutant which has as its new N-terminal residue a glycine that corresponds to G74 of the full-length human GalNAcT2 set forth in SEQ ID NO:2, and Δ94, a truncation mutant which has as its new N-terminal residue a glycine that corresponds to G95 of the full-length human GalNAcT2 set forth in SEQ ID NO:2.

[0199] Truncated human polypeptide N-acetylgalactosaminyltransferase II (GalNAcT2) was expressed as maltose binding protein (MBP)-fusion proteins in inclusion bodies from *E. coli* JM109 cells. The production of active enzyme was examined by refolding and assaying against two polypeptide acceptors. Therefore, described herein is the generation of several truncated forms of human polypeptide GalNAcT2 as maltose binding protein fusion proteins in *E.coli* JM109 cells. The recombinant proteins are refolded from isolated inclusion bodies using the Hampton FoldIt screen kit (Hampton Research, Aliso Vieja, CA). All four constructs were expressed in JM109 *E.coli* at levels of approximately 2g/L culture media.

[0200] PCR (Polymerase Chain Reaction) amplifications were performed in a final reaction volume of 50 µl containing 5 µl of template DNA (11 µg/ml, 100-fold diluted pBKS-Full ppGalNAcT2), 40 pmol of 5'- primer and 3'- primer, 10 nmol of dNTP mixture, and 5 units of Herculase™ Enhanced DNA Polymerase under the conditions of 31 cycles of denaturation at 95°C for 45 seconds, annealing at 62°C for 45 seconds, and extension at 74°C for 170 seconds. PCR products were subjected to 1% agarose gel electrophoresis. DNA fragments

were excised and purified by QIAEX II gel extraction kit (Qiagen, Valencia, CA). Table 1 illustrates the primers used in the PCR reactions.

Table 1: Primers used in cloning ppGalNAcT2

Sense Primers:

For N41R (relates to Δ40):
5' CGCGGATCCAGGAAGGAGGACTGGAATG 3' (SEQ ID NO:11)
BamHI
For N52K (relates to Δ51):
5' CGCGGATCCAAAAAGAAAGACCTTCATCACAGC 3' (SEQ ID NO:12)
BamHI
For N74G (relates to Δ73):
5' CGCGGATCCGGGAAAGTACGGTGGCCAGAC 3' (SEQ ID NO:13)
BamHI
For N95G (relates to Δ94):
5' CGCGGATCCGGGCAGGACCCTTACGCC 3' (SEQ ID NO:14)
BamHI

Antisense Primer with STOP codon:

5'-CTGCTCGAGCTACTGCTGCAGGTTGAGCG 3' (SEQ ID NO:15)
XhoI Stop

[0201] Gel-purified PCR products were digested with *BamH*I and *Xho*I, gel purified again and ligated into a pCWin2MBP vector previously digested by the same restriction enzymes. The ligated products were transformed into *E. coli* DH5α electrocompetent cells. The transformants were plated on LB Agar plates with 50 μg/ml Kanamycin and incubated at 37°C overnight. Three colonies were picked for each construct and cultured in LB medium containing 15 μg/ml kanamycin. Plasmid DNAs were purified by QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA) and screened by restriction mapping with *BamH*I and *Xho*I. The plasmids having the correct digest patterns were transformed into JM109 chemical competent cell.

[0202] JM109 cells were cultured in a 15 ml culture tube containing 6 ml LB medium and 15 μg/ml of kanamycin overnight at 37°C with rapid shaking (250 rpm). For each culture, two milliliters of starting culture was transferred to a 50 ml centrifuge tube containing 23 ml LB medium with 15 μg/ml kanamycin and incubated at 37°C with rapid shaking for 3 hours. Isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to a final concentration of 0.4 mM to induce the protein expression. After shaking at 37°C (250 rpm) for another 3 hours, cells were harvested by centrifugation at 3,500 x g for 10 minutes. The cell pellets were then

resuspended in 0.6 ml of 20 mM Tris-HCl buffer (pH 8.5) containing 1% Triton X-100. Lysozyme (100 μg) and DNaseI (2 μg) were then added. The mixture was shaken at 37°C in an incubator shaker for 45 minutes before being transferred to a 1.5 ml microcentrifuge tube. Lysate was separated from inclusion bodies (IB) by centrifugation at 14,000 rpm for 5 minutes.

[0203] Each sample for SDS-PAGE separation was prepared by mixing 5 μl of whole cells suspension, lysate, or inclusion bodies suspension with 5 μl of 2 x Tris-Glycine SDS sample buffer and 1.1 μl of DTT (1 M). The mixture was heated at 98°C for 5 minutes, cooled to room temperature, and loaded to each well of a 1.0 mm x 15 well 4-20% Tris-Glycine gradient gel. The electrophoresis was conducted at 120 V for 100 minutes. The gel was then stained for 2 hours and de-stained with distilled water (see Figures 4-6).

[0204] Inclusion bodies were dissolved at 20 mg/ml (high protein concentration) or 2 mg/ml concentration (low protein concentration) in solubilizing buffer containing 4 M Guanidine-HCl, 100 mM Tris-HCl, pH 9.0, 5 mM EDTA, and 10 mM DTT. Refolding of inclusion bodies by Hampton FoldIt Screen Kit was carried out by following the manufacturer's protocol, except that a 10-fold less volume was used (100 μl -scale) (Hampton Products, Aliso Viejo, CA).

[0205] Non-radioactive enzyme activity assays for lysates were carried out in a 0.5 μl microcentrifuge tube at 37°C for overnight in a final volume of 10 μl containing 50 mM MES buffer, pH 6.0, $MnCl_2$ (15 mM), $MgCl_2$ (15 mM), NaCl (0.15 M), UDP-GalNAc (5 mM), 1.5 μg G-CSF (acceptor), and 2.15 μl of lysate sample. Enzyme was substituted by H_2O as a negative control. Purified recombinant ppGalNAcT2 (0.5 μl) from Sf9 baculovirus expression system was used as the positive control. An assay for refolded inclusion bodies was performed in a similar manner as described for the lysates, except that Interferon α2b (4 μg) was used as the acceptor for the enzyme and the volume of the sample added to the reaction mixture was 5.65 μl.

[0206] DNA fragments for ppGalNAcT2 genes (about 1.5 kb) were successfully amplified by PCR as shown in Figure 5. Vector plasmid DNA pCWin2MBP was digested by BamHI and XhoI, and purified on a 1% agarose gel. The gel purified DNA fragment was digested by the same two enzymes and purified. After digestion, the DNA fragments were clean as visualized on an agarose gel (Figure 2B).

[0207] BamHI and XhoI digestion of the plasmids purified from the selected twelve colonies showed predicted correct pattern on a 1% agarose gel. The size of the vector was around 6.2 kb, and the inserts were approximately 1.5 kb. Maltose-binding protein (MBP) expressed in the JM109 transformed with pCWin2MBP vector plasmid showed a band at around 43 kDa. Over 90% of the proteins in the whole cells are MBP. The #2 colony of the construct N41R expressed a shorter protein than expected, indicating the occurrence of mutation. All other eleven colonies showed a band at about 100 kDa for MBP-ppGalNAcT2 fusion proteins, with over 80% of the total proteins were the target fusion proteins.

[0208] Gel electrophoresis showed that most of the MBP was expressed as a soluble form in cell lysate (Figures 1 and 2). The overexpressed protein in the wrong construct (colony #2) for N41R was also observed in the cell lysate. However, most of the MBP-ppGalNAcT2 fusion proteins were in inclusion bodies (Lane 1 and 3 – 12 in Figures 1 and 2). Over 90% of the proteins in the inclusion bodies were the MBP fusion proteins of interest.

[0209] In summary, four truncated forms of human polypeptide GalNAcT2 were successfully cloned into pCWin2MBP vector and expressed in *E. coli* JM109 as MBP fusion proteins in inclusion bodies. The level of expression of enzyme in inclusion bodies was about 2 g/L. As estimated from the SDS-PAGE, over 80% of the inclusion bodies were the target MBP-ppGalNAcT2 fusion proteins.

Example 2: Development of Protein Refolding Conditions for E. Coli Expressed MBP-Human GalNAcT2

[0210] Refolding experiments on MBP-GalNAcT2 were carried out on a 1 ml scale, with four different MBP-GalNAcT2 DNA constructs and under 16 different possible refolding conditions. Refolding was performed using the Hampton Research Foldit kit (Hampton Research, Aliso Viejo, CA) and the assays were performed via radioactive detection of [^{3}H] UDP-GalNAc addition to a MuC-2 peptide and via matrix-assisted laser desorption ionization mass spectrometry (MALDI) analysis utilizing addition of GalNAc to Interferon α-2b and G-CSF. The data illustrates that E.coli-expressed MBP-GalNAcT2 can be refolded into an active enzyme. It appears that under refolding conditions 8 and 15, found in Hampton Research's Foldit kit (Hampton Research, Aliso Viejo, CA), active conformations of MBP-GalNAcT2, construct 1 and 2, were identified. Success was indicated by the [^{3}H] UDP-GalNAc assay and later confirmed by interferon α-2b (IFα-2b) and granulocyte-colony

stimulating factor (G-CSF) -based glycosyltransferase assays. The specific methods and data of this study are presented herein.

[0211] As described elsewhere herein, GalNAcT2 constructs used in the present invention comprised DNA encoding various amino terminal amino acid truncation mutants of the original human GalNAcT2 protein, including the following constructs, which begin with the N-terminal amino acid as indicated:

Construct 1 – pCWin2 MBP-GalNAcT2 – R41 Arginine (924aa, 103682.5MW),
Construct 2 – pCWin2 MBP-GalNAcT2 – K52 Lysine (913aa, 102286.0MW),
Construct 3 – pCWin2 MBP-GalNAcT2 – G74 Glycine (891aa, 99799.3MW), and
Construct 4 – pCWin2 MBP-GalNAcT2 – G95 Glycine (870aa, 97419.8MW) .

[0212] Constructs were first expanded to 2 ml starter cultures by inoculating 2 ml of Martone L-Broth containing 10µg/ml Kanamycin sulfate with a pipette tip scraping from the particular glycerol stock culture. This procedure was performed on all four constructs for a total of four starter cultures. Starter cultures were incubated overnight at 37°C, with rotary shaking at 250rpm. From the overnight cultures, four 275 ml Martone L-Broth cultures containing 10µg/ml Kanamycin sulfate were prepared. Each of these cultures was inoculated with 275µL of one of the 2 ml starter cultures of constructs 1 through 4. These 275 ml cultures were incubated overnight at 37°C, with shaking at 250rpm.

[0213] Lastly, four 1L Martone L-Broth cultures containing 10µg/ml Kanamycin sulfate were prepared. Each of these cultures was inoculated with 40 ml of one of the 275 ml cultures of constructs 1 though 4. These 1L cultures were incubated at 37°C, with shaking at 250rpm, until the OD600 measured approximately 1.0. Upon reaching this point, IPTG was added to each of the four 1L cultures to a final concentration of 0.4mM. Cultures were then allowed incubate overnight at 37°C, with shaking at 250rpm.

[0214] One-liter cultures containing JM109 pCWin2 MBP-GalNAcT2 constructs, designated numbers 1 through 4, were transferred to 1L centrifuge bottles. Cultures were then centrifuged at 5000rpm for 30 minutes at 4°C. Supernatants were removed and the pellets were weighed. The pellets from each sample were then washed to isolate the inclusion bodies (IBs). The pellet of each construct was first resuspended in 15 ml of 20mM Tris-HCl pH=8.5, 5mM EDTA and then lysed by two passages through a french press at 12,000psi.

[0215] The lysates for each construct were then centrifuged at 5000rpm, 25°C for 5 minutes in 50 ml disposable tubes. The supernatants were removed and the pellets were resuspended in 25 ml of 20mM Tris-HCl pH=8.5, 1% Triton X-100. The suspensions

incubated at room temperature for 10 minutes. The suspensions were then centrifuged at 5000 rpm, 25°C for 5 minutes. The supernatants were then removed and the samples were resuspended for a second time in 25 ml of 20mM Tris-HCl pH=8.5, 1% Triton X-100 and allowed to incubate at room temperature for 10 minutes. The suspensions were again centrifuged at 5000 rpm, 25°C for 5 minutes. The supernatants were removed and a third wash was performed by resuspending the pellets in 25 ml of 20mM Tris-HCl pH=8.5, 1% Triton X-100. The suspensions sat at room temperature for 10 minutes and then were centrifuged at 5000 rpm, 25°C for 5 minutes. The supernatants from each sample were removed and the pellets were weighed. The pellets were then diluted to 20mg/ml by resuspending them in the appropriate volume of 20mM Tris-HCl pH=8.5, 5mM EDTA. One-ml aliquots were made from these suspensions for each of the four constructs and stored at -20°C. These aliquots represent the triple washed IBs or "TWIBs."

[0216] Solubilization buffer was prepared with the following constituents: 6M Guanidine HCl, 5mM EDTA, 50mM Tris-HCl pH=8 and 10mM DTT. 1 ml of this solution was added to a 20mg aliquot of TWIBs to yield a 20mg/ml solution. The solution was incubated overnight on the bench top to solubilize IBs. This procedure was performed on a TWIB aliquot of each MBP-GalNAcT2 construct to provide protein for refolding experiments.

[0217] To screen refolding conditions that may result in an active form of E.coli expressed MBP-GalNAcT2, a Hampton Foldit Screening kit was utilized (Hampton Products, Aliso Viejo, CA). The composition of each of the refold buffers is found in Table 2.

Table 2: Refold Conditions from Hampton Research Foldit kit (Hampton Research, Aliso Viejo, CA)

Refold Matrix															
Refold Condition	55mM Tris	55mM MES	264mM NaCl /	10.56mM NaCl /	0.055% PEG	550mM Guanidine	1.1mM EDTA	2.2mM MgCl2 /	440mM Sucrose /	440mM Sucrose	550mM L-Arginine	1mM DTT	1mM GSH	0.1mM GSSG	Protein Concentration
1	X		X		X		X					X			0.1
2		X		X		X		X					X	X	0.1
3		X		X	X	X	X		X				X	X	0.1
4	X		X					X	X			X			0.1
5		X	X					X		X			X	X	1
6	X			X	X	X	X			X		X			1
7	X			X		X		X			X	X			1
8		X	X		X		X				X		X	X	1
9		X	X		X	X		X		X		X			0.1
10	X			X			X			X			X	X	0.1
11	X			X	X			X			X		X	X	0.1
12		X	X			X	X				X	X			0.1
13	X		X			X	X						X	X	1
14		X		X	X			X				X			1
15		X		X			X		X			X			1
16	X		X		X	X		X	X				X	X	1

[0218] For a given refold condition, 950μL of refold buffer was combined with 50μL of solubilized protein (for high protein concentration conditions) or 995μL of refold buffer was combined with 5μL of solubilized protein (for low protein concentration conditions). Refolding reactions were placed on a rotary shaker in the cold room (4°C) overnight.

[0219] From results obtained in the screen, it was determined that refold conditions 3, 8, 11, 12, 15 and 16 yielded the most promising results for constructs 1 and 2. Additional refolding reactions were performed with under those conditions using G-50 gel filtration instead of dialysis to yield more concentrated protein refold samples (See Refold Purification section for methods). From those experiments, further refinement was achieved and conditions 8 and 11 were found to be optimal. More specifically, condition 15 was optimal in an overnight incubation rotating and condition 8 was found to be optimal remaining still in a 5 day incubation.

[0220] Protein refold samples were first purified by dialysis against 20mM Tris-HCl, pH=8.5. 100μL of each refold sample was dialyzed. Dialysis was conducted in a beaker containing 20mM Tris-HCl pH=8.5 with slow stirring. Samples were placed at 4°C and allowed to dialyze overnight. Resulting retentate was used in a radioactive activity assay, as discussed elsewhere herein. As an alternative method to yield more concentrated protein samples, MBP-GalNAcT2 refold samples were purified by use of G-50 Macro Spin Columns (Harvard Bioscience, Holliston, MA). Caps were removed from the G-50 columns and columns were placed into 2 ml microcentrifuge tubes. H_2O (500 μl) was added to each column and the columns were allowed to incubate for 15 minutes to hydrate. The columns were then centrifuged at ~2000 x g for 4 minutes after which they were transferred to new 2 ml centrifuge tubes. Each refold solution (150μl) was applied to one of the columns. Columns were then centrifuged at 2000 x g for ~2 minutes. Resulting permeates represented the purified refold samples.

[0221] A radiolabeled [^{3}H]-UDP-GalNAc assay was performed to determine the activity of the E.coli-expressed refolded MBP-GalNAcT2 by monitoring the addition of radiolabeled GalNAc to a peptide acceptor. The acceptor was a MuC-2 – like peptide having the sequence MVTPTPTPTC (SEQ ID NO:16). The peptide was dissolved in 1M Tris-HCl pH=8.0. The initial screen was performed on refolded protein samples which had been purified by dialysis. Subsequent refold samples were freshly refolded and purified by G-50 gel filtration. The assay included protein refold samples, GalNAcT2 from Baculovirus as a positive control, a

negative control sample with all the components except enzyme and a maximum input sample which contained all components except enzyme. A total of 19 samples were tested. The assay solution consisted of the components listed in Table 3:

Table 3: GalNAcT2 assay reaction composition.

Component	Dilution	Volume (µl)	Final Concentration
0.25M Tris-HCl	N/A	5	25mM
2.5% Triton X-100	N/A	5	0.25%
100mM $MnCl_2$	N/A	5	10mM
$[H^3]$ UDP-GalNAc 0.1mCi/ml	0.5µl in 4.5µl	5	50nCi
1mM UDP-GalNAc	N/A	5	0.1mM
10mM MuC2 Peptide	0.5µl in 4.5µl	5	0.1mM
Enzyme		20	

[0222] For each of the refold samples, 30µL of the reaction mixture were combined with 20µL of the refold sample. For the negative control, 20µL H_2O was combined with 30µL of the reaction mixture. For the positive control, 1µL of GalNAcT2 Baculovirus enzyme was added in addition to 19µL of H_2O to form a 30µL reaction mixture. For the "maximum input" sample, 30µL of the reaction mixture was combined with 20µL of dH_2O. Reactions were incubated at 37°C for 30 minutes. 100 ml DOWEX AG 1X8 (chloride form) was washed by combining 100 ml of resin and 100 ml of H_2O and mixing well. The water was poured off the resin and another 100 ml of H_2O was added, mixed and removed. The resin was resuspended one final time in 100 ml of dH_2O. After the GalNAcT2 assay reaction had incubated for 30 minutes, 1 ml of resuspended resin in H_2O was added to each reaction (except for the maximum input sample). Samples were vortexed briefly and then loaded into filter columns and allowed to drain by gravity into scintillation vials. 5 ml of scintillation solution was added to each of the samples and standards. Samples were shaken briefly and loaded on the scintillation counter and radioactivity measured.

[0223] An IFα-2b assay was performed to determine whether E.coli-expressed refolded MBP-GalNAcT2 could transfer GalNAc to an interferon α-2b acceptor from a UDP-GalNAc donor. From data obtained in the refold screen (see the $[^3H]$UDP-GalNAcT2 assay description elsewhere herein), it was shown that MBP-GalNAc constructs 1 and 2 in refold buffers 8 and 15 yielded the most active enzymes, as determined by the radioactive assay. Therefore, in the IFα-2b assay, constructs 1 and 2 in refold buffers 8 and 15 were assayed for transferase activity. Additionally, as a positive control, GalNAcT2 from a Baculovirus system was assayed as well.

[0224] The assay consisted of reaction buffer (27mM MES, pH=7, 200mM NaCl, 20mM MgCl2, 20mM MnCl2, and 0.1% Tween 80), IFα-2b Protein (2mg/ml in 50mM MES pH=6, 150mM NaCl, 0.05% Tween 80, 0.05% NaN_3), and 100mM UDP-GalNAc. The assay solution was prepared as shown in Table 4 for each reaction.

Table 4: Parameters for IFα-2b acceptor GalNAcT2 activity assay

	Reaction Components	Reaction Component Volumes	Final Concentration
Reaction Buffer	MES, pH=7		20mM
	NaCl	5µl from Rxn Buffer (additional concentration from IFα-2b dilution buffer)	150mM
	$MgCl_2$		5mM
	$MnCl_2$		5mM
	Tween 80		0.05%
	2mg/ml IFα-2b Protein	10µl	1mg/ml
	100mM UDP-GalNAc	0.6µl	3mM

[0225] For each refold sample, 4.4µL of sample were added to 15µL of reaction solution. For the positive control, 1µL of standard GalNAcT2 Baculovirus was added along with 3.4µL of H_2O to one tube. Reactions were incubated at 32°C on a rotary shaker for several days, during which time an overnight time point and a 5 day time point were assayed by MALDI.

[0226] The above assay was performed to determine whether E.coli-expressed refolded MBP-GalNAcT2 could transfer GalNAc to G-CSF acceptor from a UDP-GalNAc donor. As above, construct 2 in refold buffer 8 was assayed for GalNAcT2 activity. Additionally, as a positive control, GalNAcT2 from Baculovirus was assayed. The assay consisted of reaction buffer (27mM MES, pH=7, 200mM NaCl, 20mM MgCl2, 20mM MnCl2, and 0.1% Tween 80), G-CSF Protein (2mg/ml in H_2O), and 100mM UDP-GalNAc. The assay solution was prepared for each reaction as shown in Table 5.

Table 5: Parameters for G-CSF acceptor GalNAcT2 activity assay

	Reaction Components	Reaction Component Volumes	Final Concentration
Reaction Buffer	MES, pH=7	5µl of Rxn Buffer	20mM
	NaCl		150mM
	$MgCl_2$		5mM
	$MnCl_2$		5mM
	Tween 80		0.05%
	2mg/ml G-CSF	10µl	1mg/ml
	100mM UDP-GalNAc	0.6µl	3mM

[0227] For the refold sample, 4.4µL of sample were added to 15µL of reaction solution. For the positive control, 1µL of standard GalNAcT2 Baculovirus was added along with 3.4µL of H_2O to one tube. Reactions were incubated at 32°C on a rotary shaker for 4 days, at the end of which a sample was taken and assayed by MALDI.

[0228] Pellet weights and inclusion body weight were determined for each of the four 1L JM109 pCWin2 MBP-GalNAcT2 constructs 1 through 4 cultures, as shown in Table 22.

Table 6: Cell pellet weights versus inclusion body weights

Pellet and Inclusion Body Weights from 1L JM109 pCWin2 MBP-GalNAcT2 Cultures		
JM109 pCWin2 MBP-GalNAcT2 Construct	Cell Pellet Weight (g)	Inclusion Body Weight (g)
1	5.04	2.04
2	5.24	2.19
3	4.89	2.42
4	4.30	2.44

[0229] The expression of MBP-GalNAcT2 was observed by way of the SDS-Page gel analysis of JM109 pCWin2 MBP-GalNAcT2 whole cell samples before and after induction by IPTG (Figure 7). The protein gel shows a clear increase in protein expression in the induced state compared to the uninduced state. Furthermore there is a distinct band at ~100kDa that substantially increases after induction which correlates to the expected size of the MBP-GalNAcT2 band.

[0230] Protein samples were diluted by combining 950µL of H_2O with 50µL of protein sample. Samples were then analyzed using a UV spectrophotometer. Protein concentration

was calculated from absorption values and the molar extinction coefficients: Construct 1 – 0.65mg/ml per 1 A_{280} unit, Construct 2 – 0.64mg/ml per 1 A_{280} unit, as shown in Table 7.

Table 7: Protein concentration of 1L JM109 pCWin2 MBP-GalNAcT2 Cultures after Solubilization and G-50 Purification

JM109 pCWin2 MBP-GalNAcT2 Construct	A_{280} After Solubilization	Protein Concentration (mg/ml)	A_{280} After G-50 Purification	Protein Concentration (mg/ml)
1	0.2827	2.5	0.0100	0.156
2	0.2531	2.4	0.0160	0.102

[0231] Inclusion bodies obtained from JM109 pCWin2 MBP-GalNAcT2 constructs 1 and 2 were analyzed using SDS-PAGE to verify the presence of MBP-GalNAcT2. The protein was clearly observed in both lanes of the gel, running at approximately 100kDa (Figure 8).

[0232] All four constructs were tested in a [^{3}H]UDP-GalNAcT2 assay under all 16 refold conditions available in the Hampton Foldit kit (Hampton Research, Aliso Viejo, CA). Refolded truncated enzymes were purified by dialysis and then tested for activity using the radioactive assay, as shown in Table 8.

Table 8: Results of the GalNAcT2 activity assay for refolded proteins

	Raw CPM															
Refold Condition	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Colony 1	112	119	155	150	131	168	167	243	111	144	218	218	166	114	214	194
Colony 2	119	121	251	143	132	156	160	221	121	166	230	184	139	137	224	222
Colony 3	125	113	207	139	96	123	143	170	100	110	134	184	143	114	174	180
Colony 4	122	123	125	150	132	120	119	135	127	121	148	154	121	95	157	165
Average for Refold Condition	119.5	119.0	184.5	145.5	122.8	141.8	147.3	192.3	114.8	135.3	182.5	185.0	142.3	115.0	192.3	190.3
Negative Control		102														
Positive Control		1585														
	Corrected CPM															
Refold Condition	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Colony 1	10	17	53	48	29	66	65	**141**	9	42	**116**	**116**	64	12	**112**	92
Colony 2	17	19	**149**	41	30	54	58	**119**	19	64	**128**	82	37	35	**122**	**120**
Colony 3	23	11	105	37	-6	21	41	68	-2	8	32	82	41	12	72	78
Colony 4	20	21	23	48	30	18	17	33	25	19	46	52	19	-7	55	63
Average for Refold Condition	17.5	17.0	**82.5**	43.5	20.8	39.8	45.3	**90.3**	12.8	33.3	**80.5**	**83.0**	40.3	13.0	**90.3**	**88.3**

[0233] Results from this assay indicated that refold conditions 3, 8, 11, 12, 15 and 16 provided the highest CPM and therefore the greatest potential GalNAcT2 activity. Furthermore it appeared that construct 2 yielded the greatest number of positive hits in this assay, therefore efforts were focused on this construct.

Table 9: Results from focused overnight refold of truncated enzymes

	Raw CMP3					
Refold Condition	**3**	**8**	**11**	**12**	**15**	**16**
Construct 2	924	1197	689	1585	1701	1561
Negative Control		277				
Positive Control		4919	(1µl of 200ug/ml STD Enzyme)			

	Corrected CPM3					
Refold Condition	**3**	**8**	**11**	**12**	**15**	**16**
Construct 2	647	920	412	1308	1424	1284

Activity:

$$U/L = \frac{CMP \times (nmoles\ Donor) \times 100\mu l/ml}{(Input\ CPM) \times (0.35/0.55) \times (Assay\ Incubation\ Time(minutes)) \times Volume\ Enzyme\ (\mu l)}$$

nmoles Donor (UDP-GalNAc)	5
Assay Incubation Time (minutes)	30
Volume Enzyme (µl)	20
Maximum Input	48998

	Activity U/L						
	3	**8**	**11**	**12**	**15**	**16**	**Positive Control**
Construct 2	0.17	0.25	0.11	0.35	0.38	0.34	26.29

[0234] In this assay, construct 2 was tested under refold conditions 3, 8, 11, 12, 15 and 16 from the Hampton Foldit kit (Hampton Research, Aliso Viejo, CA). These refolded enzymes were purified by G-50 gel filtration and then tested for activity by the radioactive assay. Results indicate that after overnight incubation on a rotator, greatest activity was obtained from refold condition 15.

Table 10: GalNAcT2 activity results from 5 day refolding experiment

	Raw CMP3 - 5 Day Refold					
Refold Condition	**3**	**8**	**11**	**12**	**15**	**16**
Construct 2	185	2288	186	226	496	270
Negative Control		129				
Positive Control		3612	(1µl of 200ug/ml STD Enzyme)			

	Corrected CPM3 - 5 Day Refold					
Refold Condition	**3**	**8**	**11**	**12**	**15**	**16**
Construct 2	56	2159	57	97	367	141

Activity:

$$U/L = \frac{CMP \times (nmoles\ Donor) \times 100\mu l/ml}{(Input\ CPM) \times (0.35/0.55) \times (Assay\ Incubation\ Time(minutes)) \times Volume\ Enzyme\ (\mu l)}$$

nmoles Donor (UDP-GalNAc)	5
Assay Incubation Time (minutes)	30
Volume Enzyme (µl)	20
Maximum Input	47527

	Activity U/L - 5 Day Refold						
	3	**8**	**11**	**12**	**15**	**16**	**Positive Control**
Construct 2	0.02	0.59	0.02	0.03	0.10	0.04	19.90

[0235] In this assay, construct 2 was tested under refold conditions 3, 8, 11, 12, 15 and 16 from the Hampton Foldit kit (Hampton Research, Aliso Viejo, CA) after being rotated overnight at 4°C and left resting at 4°C for 5 days. These refolded enzymes were purified by G-50 gel filtration and then tested for activity by the radioactive assay. Results indicated that after 5 days in refold buffer 8, construct 8 displayed the highest activity. Therefore it was determined that conditions 8 and 15 had the greatest potential for producing a properly folded and active MBP-GalNAcT2.

[0236] An IFα-2b assay was performed on overnight refolds of constructs 1 and 2 in refold buffer 15 (1-15 and 2-15, respectively) and was incubated at 32°C for 5 days. Time points were taken of the IFα-2b reaction at 16 hours and 5 days. The results indicate that the parental peak for IFα-2b is at MW ~19267. A successful reaction would be indicated by addition of ~203 molecular weight to that peak. From the 5 day data for refolds 1-15 and 2-15, a developing peak was observed at ~119478 and ~19473 respectively, a difference of approximately 203 MW. This data illustrated that GalNAc was added to IFα-2b by the refolded GalNAcT2 protein, thereby confirming the activity that was reported elsewhere herein by the radioactive assay.

[0237] Additionally, the IFα-2b assay was performed with the 5-day refolded enzymes of constructs 1 and 2 in refold buffer 8 (1-8 and 2-8, respectively). The IFα-2b reactions were again allowed to incubate at 32°C for 3 days. Reactions were analyzed at the 3 day time point. The results indicated that the parental peak for IFα-2b is at MW ~19263. A successful reaction would be indicated by the addition of ~203 molecular weight to that peak. From the 3 day data for refolds 1-8 and 2-8 a developing peak is seen at ~19462 and 19469 respectively, again a difference of approximately 203 MW. This data again indicated that GalNAc was added to IFα-2b by the refolded GalNAcT2 protein and confirmed what was reported by the radioactive assay.

[0238] A G-CSF assay was performed on the 5-day refolded enzymes of construct 2 in refold buffer 8. The G-CSF reaction was allowed to incubate at 32°C for 4 days. The reaction was analyzed at the 4 day time point. The parental peak for G-CSF is expected at MW ~18786. A successful reaction would be indicated by addition of ~203 molecular weight to that peak. From the 3 day data for refolded enzymes 2-8, a developing peak was observed at ~19001, a difference of approximately 203 MW. This data again indicated that GalNAc was added to G-CSF by the refolded GalNAcT2 protein and confirmed what was reported by the radioactive assay and the IFα-2b assay as reported elsewhere herein.

[0239] In summary, the data presented herein illustrates that E.coli-expressed MBP-GalNAcT2 can be refolded into an active enzyme. Under refold conditions 8 and 15, found in Hampton Research's Foldit kit (Hampton Research, Aliso Viejo, CA), active conformations of MBP-GalNAcT2 construct 1 and 2 were obtained. The generation of a functional refolded protein was shown using radioactive, IFα-2b and G-CSF assays, which demonstrated the transfer of GalNAc to a polypeptide by GalNAcT2 truncation mutants of the present invention.

[0240] As discussed elsewhere herein, GalNAcT2 truncation mutants of the present invention are also useful for the transfer of a glycosyl-polyethyleneglycol ("glycosyl-PEG") conjugate to a polypeptide, also known as "glycoPEGylation" of a polypeptide. Using a purified, refolded Δ51 GalNAcT2-MBP fusion made according to the present invention, it was shown that Δ51 GalNAcT2-MBP is capable of transferring a GalNAc-sialic acid (SA)-PEG conjugate to G-CSF.

[0241] A glycoPEGylation reaction mixture was prepared in order to glycoPEGylate G-CSF. The reaction mixture contained 5 μl of Δ51 GalNAcT2-MBP (20 μU), 2 μl of GalNAc-

α2,6-sialyltransferase (ST6GalNAcI), 6.25 mM $MnCl_2$, 15 mM UDP-GalNAc, 0.75 mM CMP-SA-PEG (20K), and between 2 µl and 10 µl of 2 mg/ml G-CSF. Gel electrophoresis of the reaction products demonstrated that Δ51 GalNAcT2-MBP transferred a GalNAc-sialic acid (SA)-PEG conjugate to G-CSF (Figure 9).

Example 3: Optimization of Purification and Refolding of Δ51 GalNAcT2-MBP

[0242] Δ51 GalNAcT2 refolding and purification development as set forth herein demonstrates the utility of a two column purification procedure for purification of GalNAcT2 mutants. The use of Q Sepharose Fast Flow in binding mode and Q Sepharose XL in binding and flow through mode as an initial purification step has been explored. Q Sepharose XL in flow through mode using a NaCl concentration of 100mM in the load led to best recovery and purity of active Δ51 GalNAcT2-MBP. The use of Hydroxyapatite Type I has been considered as a second column step. Initial data indicate Δ51 GalNAcT2-MBP binds to this resin and can be eluted as an active enzyme with a phosphate gradient.

[0243] Δ51 GalNAcT2-MBP was cloned and expressed as set forth elsewhere herein. To produce double-washed inclusion bodies (DWIBs) containing the expressed Δ51 GalNAcT2-MBP, harvested cell pellet was resuspended in 10mM Tris/ 5mM EDTA pH 7.5 (5mL/g cells) and lysed in two passes using a microfluidizer at 12,000psi. Inclusion bodies were harvested by centrifugation at 6,000 rpm for 20 min in a Sorvall RC-3B. The pellet was washed twice by resuspension in above buffer at 5mL/g pellet followed by centrifugation at 6,000 RPM for 20min. DWIBs were aliquoted and stored at -20°C.

[0244] Initial studies indicated that urea solubilization leads to higher Δ51 GalNAcT2-MBP activities of refolded material than does guanidine hydrochloride solubilization. Therefore, Δ51 GalNAcT2-MBP was solubilized in 7M urea/ 50mM Tris/ 10mM DTT/ 5mM EDTA pH 8.0 for all subsequent experiments.

1. Refolding experiments – pH scout

[0245] A pH scout was performed to identify the best pH for Δ51 GalNAcT2-MBP refolding.

Table 11: Reaction parameters for pH scouting of Δ51 GalNAcT2-MBP refolding conditions

Sample ref. no.:	1b	1a	2a	2b	3
MES (mM)	50	50			
Tris (mM)			50	50	50
L-Arginine (mM)	550	550	550	550	550
NaCl (mM)	250	250	250	250	10
KCl (mM)	10	10	10	10	
PEG 3350 (%)	0.05	0.05	0.05	0.05	0.05
L-cysteine (mM)	4	4	4	4	4
L-cystamine dihydrochloride (mM)	1	1	1	1	1
$MnCl_2$ (mM)					1
pH	5.5	6.5	8.0	8.5	8.0

[0246] Δ51 GalNAcT2-MBP refolds were performed by solubilizing 2.5g of DWIB's in 250 mL of 7M urea/ 50mM Tris/ 10mM DTT/ 5mM EDTA pH 8.0 at 4°C. 50mL solubilized Δ51 GalNAcT2-MBP DWIB's were added to 1L of refold buffer at 4°C while stirring (21-fold dilution – 0.5mg/mL). Refolding was allowed to proceed for 20.5h at 4°C with stirring.

[0247] Refolds were filtered using a Cuno Zeta Plus BioCap (Cuno, Meriden, CT), concentrated 4-fold and diafiltered on a 1 ft2 30kDa MWCO TFF (regenerated cellulose) filter at constant volume with 5 diavolumes of 10mM Tris/ 5mM NaCl pH 8.

[0248] Concentrated and diafiltered refolds were loaded onto a pre-equilibrated 48mL Q Sepharose Fast Flow column (Amersham Biosciences, Piscataway, NJ) and washed with 2 column volumes (CVs) of low salt buffer (10mM Tris/5mM NaCl pH 8.0). Protein was eluted with a 15CV gradient from 0 to 50% high salt buffer (10mM Tris/1M NaCl pH 8.0) followed by a 1CV gradient to 100% high salt buffer. The column was regenerated with 0.5M NaOH.

[0249] The highest Δ51 GalNAcT2-MBP activity was achieved using refold 2a conditions (pH 8.0) in combination with urea solubilization. Active Δ51 GalNAcT2-MBP eluted early during QSFF elution. The 1L refold yielded a total of 420mU Δ51 GalNAcT2-MBP.

[0250] Additional refolding conditions for Δ51 GalNAcT2-MBP were screened. Refolding buffer containing 55 mM MES pH 6.5, 264 mM NaCl, 11 mM KCl, 0.055% PEG 3350 and 550 mM L-Arginine and refolding buffer containing 55 mM Tris-HCl pH 8.0, 10.56 mM NaCl, 0.44 mM KCl, 0.055% PEG 3350 and 550 mM L-arginine were screened. Four conditions were screened using the two buffers, namely, solubilization at pH 6.5 followed by refolding at pH 6.5, solubilization at pH 6.5 followed by refolding at pH 8.0, solubilization at

pH 8.0 followed by refolding at pH 6.5, and solubilization at pH 8.0 followed by refolding at pH 8.0. Assays of Δ51 GalNAcT2-MBP refolded under all four conditions demonsrated enzymatic activity, the ability to transfer GalNAc to GCSF.

2. Δ51 GalNAcT2-MBP Purification

[0251] The use of Q Sepharose Fast Flow (QSFF) and Q Sepharose XL (QXL) (Amersham Biosciences, Piscataway, NJ) in Δ51 GalNAcT2-MBP purification was examined. QSFF was used in binding mode. For this purpose, concentrated diafiltered Δ51 GalNAcT2-MBP refolds (in 10mM Tris/5mM NaCl pH 8.0 - A) were applied onto a pre-equilibrated 50mL QSFF column and eluted using a gradient from 10mM Tris/ 5mM NaCl pH 8.0 to 50% 10mM Tris/ 1M NaCl pH 8.0 (B) over 15 CV, followed by a second gradient from 50 to 100% B over 1CV.

[0252] QXL was used in binding and in flow through mode. The NaCl concentration in the concentrated diafiltered Δ51 GalNAcT2-MBP refold material (40mL each = 160mL refold volume) was adjusted to 5, 50, 100, and 200mM NaCl prior to application onto a 3.9mL QXL column. The column was washed with 2CV and bound protein was eluted with a 30CV gradient from A to B.

[0253] Δ51 GalNAcT2-MBP bound tightly to QSFF resin under above conditions with 5mM NaCl in load and equilibration buffers. Active Δ51 GalNAcT2-MBP eluted at the beginning of the major peak and appears as a doublet on a nonreduced 4-20% Tris-glycine gel. The major contaminant is a currently unidentified band running at a slightly lower molecular weight close to the 98kDa marker band. A variety of other contaminants elute with inactive Δ51 GalNAcT2-MBP in the remainder of the major peak.

[0254] Δ51 GalNAcT2-MBP bound tightly to QXL resin if the same conditions as for QSFF binding were applied (i.e. 5mM NaCl). Increasing Δ51 GalNAcT2-MBP activity was observed in flow through and wash at higher NaCl concentrations in the load. Interestingly, the major contaminating band observed in QSFF purification was not visible in the flow through if the load contained 50 and 100mM NaCl. At both NaCl concentrations the majority of active Δ51 GalNAcT2-MBP could be found in flow through and wash; only some residual Δ51 GalNAcT2-MBP activity was detected in the left shoulder of the elution peak. As observed with QSFF resin, the bulk of contaminating bands was observed in the major elution peak. Although the majority of active Δ51 GalNAcT2-MBP was located in the flow through if the salt concentration of the load was adjusted to 200mM, no significant

purification was achieved under this condition. In conclusion, optimum NaCl concentration for the use of QXL in FT mode would be higher than 50mM NaCl, but below 200mM NaCl. On the basis of these data, 100mM NaCl is a suitable concentration in the load and in the equilibration buffer in order to use the anion exchange resin in flowthrough mode.

[0255] Hydroxyapatite Type I (80μm) (BioRad, Hercules, CA) was examined as a second column step. Active Δ51 GalNAcT2-MBP partially purified over QSFF (using bind and elute mode) was used to investigate if active Δ51 GalNAcT2-MBP would bind to an HA Type I resin and would be useful to further purify the protein. For this purpose, a 2.25 mL HA Type I column was pre-equilibrated with 5mM NaPO4/ 5mM NaCl pH 7.0 (C). Active Δ51 GalNAcT2-MBP eluted from QSFF was adjusted to pH 7.0 with 1M HCl and applied onto the HA Type I column. The protein was eluted using a 20 CV gradient from 0-50% 300mM NaPO4/ 5mM NaCl pH 7.0 (D), followed by a 5 CV gradient from 50-100% D. The column was regenerated using 0.5M NaOH. The data obtained indicate that Δ51 GalNAcT2-MBP binds to hydroxyapatite type I resin and can be eluted as an active enzyme.

[0256] The disclosures of each and every patent, patent application, and publication cited herein are hereby incorporated herein by reference in their entirety.

[0257] While this invention has been disclosed with reference to specific embodiments, it is apparent that other embodiments and variations of this invention may be devised by others skilled in the art without departing from the true spirit and scope of the invention. The appended claims are intended to be construed to include all such embodiments and equivalent variations.

SEQUENCE LISTING

<110> Neose Technologies, Inc.
Johnson, Karl
Chen, Xi

<120> Truncated GalNAcT2 Polypeptides and Nucleic Acids

<130> 040853-01-5149PR

<160> 16

<170> PatentIn version 3.2

<210> 1
<211> 1713
<212> DNA
<213> human

<220>
<221> misc_feature
<223> wild-type GalNAcT2

<400> 1
atgcggcggc gctcgcggat gctgctctgc ttcgccttcc tgtgggtgct gggcatcgcc 60

tactacatgt actcgggggg cggctctgcg ctggccgggg gcgcgggcgg cggcgccggc 120

aggaaggagg actggaatga aattgacccc attaaaaaga aagaccttca tcacagcaat 180

ggagaagaga aagcacaaag catggagacc ctccctccag ggaaagtacg gtggccagac 240

tttaaccagg aagcttatgt tggagggacg atggtccgct ccgggcagga cccttacgcc 300

cgcaacaagt tcaaccaggt ggagagtgat aagcttcgaa tggacagagc catccctgac 360

acccggcatg accagtgtca gcggaagcag tggcgggtgg atctgccggc caccagcgtg 420

gtgatcacgt ttcacaatga agccaggtcg gccctactca ggaccgtggt cagcgtgctt 480

aagaaaagcc cgccccatct cataaaagaa atcatcttgg tggatgacta cagcaatgat 540

cctgaggacg gggctcttct ggggaaaatt gagaaagtgc gagttcttag aaatgatcga 600

cgagaaggcc tcatgcgctc acgggttcgg ggggccgatg ctgcccaagc caaggtcctg 660

accttcctgg acagtcactg cgagtgtaat gagcactggc tggagcccct cctggaaagg 720

gtggcggagg acaggactcg ggttgtgtca cccatcatcg atgtcattaa tatggacaac 780

tttcagtatg tgggggcatc tgctgacttg aagggcggtt ttgattggaa cttggtattc 840

aagtgggatt acatgacgcc tgagcagaga aggtcccggc aggggaaccc agtcgccct 900

ataaaaaccc ccatgattgc tggtgggctg tttgtgatgg ataagttcta ttttgaagaa 960

ctggggaagt acgacatgat gatggatgtg tggggaggag agaacctaga gatctcgttc 1020

cgcgtgtggc agtgtggtgg cagcctggag atcatcccgt gcagccgtgt gggacacgtg 1080

ttccggaagc agcacccctа cacgttcccg ggtggcagtg gcactgtctt tgcccgaaac 1140

acccgccggg cagcagaggt ctggatggat gaatacaaaa atttctatta tgcagcagtg 1200

ccttctgcta gaaacgttcc ttatggaaat attcagagca gattggagct taggaagaaa 1260

ctcagctgca agcctttcaa atggtacctt gaaaatgtct atccagagtt aagggttcca 1320

gaccatcagg atatagcttt tggggccttg cagcagggaa ctaactgcct cgacactttg 1380

ggacactttg ctgatggtgt ggttggagtt tatgaatgtc acaatgctgg gggaaaccag 1440

gaatgggcct tgacgaagga gaagtcggtg aagcacatgg atttgtgcct tactgtggtg 1500

gaccgggcac cgggctctct tataaagctg cagggctgcc gagaaaatga cagcagacag 1560

aaatgggaac agatcgaggg caactccaag ctgaggcacg tgggcagcaa cctgtgcctg 1620

gacagtcgca cggccaagag cgggggccta agcgtggagg tgtgtggccc ggccctttcg 1680

cagcagtgga agttcacgct caacctgcag cag 1713

<210> 2
<211> 571
<212> PRT
<213> human
<220>
<221> MISC_FEATURE
<223> wild-type GalNAcT2

<400> 2

Met Arg Arg Arg Ser Arg Met Leu Leu Cys Phe Ala Phe Leu Trp Val
1 5 10 15

Leu Gly Ile Ala Tyr Tyr Met Tyr Ser Gly Gly Gly Ser Ala Leu Ala
20 25 30

Gly Gly Ala Gly Gly Gly Ala Gly Arg Lys Glu Asp Trp Asn Glu Ile
35 40 45

Asp Pro Ile Lys Lys Lys Asp Leu His His Ser Asn Gly Glu Glu Lys
50 55 60

Ala Gln Ser Met Glu Thr Leu Pro Pro Gly Lys Val Arg Trp Pro Asp
65 70 75 80

Phe Asn Gln Glu Ala Tyr Val Gly Gly Thr Met Val Arg Ser Gly Gln
85 90 95

Asp Pro Tyr Ala Arg Asn Lys Phe Asn Gln Val Glu Ser Asp Lys Leu
100 105 110

Arg Met Asp Arg Ala Ile Pro Asp Thr Arg His Asp Gln Cys Gln Arg
115 120 125

Lys Gln Trp Arg Val Asp Leu Pro Ala Thr Ser Val Val Ile Thr Phe
130 135 140

His Asn Glu Ala Arg Ser Ala Leu Leu Arg Thr Val Val Ser Val Leu
145 150 155 160

Lys Lys Ser Pro Pro His Leu Ile Lys Glu Ile Ile Leu Val Asp Asp
165 170 175

Tyr Ser Asn Asp Pro Glu Asp Gly Ala Leu Leu Gly Lys Ile Glu Lys
180 185 190

Val Arg Val Leu Arg Asn Asp Arg Arg Glu Gly Leu Met Arg Ser Arg
195 200 205

Val Arg Gly Ala Asp Ala Ala Gln Ala Lys Val Leu Thr Phe Leu Asp
210 215 220

Ser His Cys Glu Cys Asn Glu His Trp Leu Glu Pro Leu Leu Glu Arg
225 230 235 240

Val Ala Glu Asp Arg Thr Arg Val Val Ser Pro Ile Ile Asp Val Ile
245 250 255

Asn Met Asp Asn Phe Gln Tyr Val Gly Ala Ser Ala Asp Leu Lys Gly
260 265 270

Gly Phe Asp Trp Asn Leu Val Phe Lys Trp Asp Tyr Met Thr Pro Glu
275 280 285

Gln Arg Arg Ser Arg Gln Gly Asn Pro Val Ala Pro Ile Lys Thr Pro
290 295 300

Met Ile Ala Gly Gly Leu Phe Val Met Asp Lys Phe Tyr Phe Glu Glu
305 310 315 320

Leu Gly Lys Tyr Asp Met Met Met Asp Val Trp Gly Gly Glu Asn Leu
325 330 335

Glu Ile Ser Phe Arg Val Trp Gln Cys Gly Gly Ser Leu Glu Ile Ile
340 345 350

Pro Cys Ser Arg Val Gly His Val Phe Arg Lys Gln His Pro Tyr Thr
355 360 365

Phe Pro Gly Gly Ser Gly Thr Val Phe Ala Arg Asn Thr Arg Arg Ala
370 375 380

Ala Glu Val Trp Met Asp Glu Tyr Lys Asn Phe Tyr Tyr Ala Ala Val
385 390 395 400

Pro Ser Ala Arg Asn Val Pro Tyr Gly Asn Ile Gln Ser Arg Leu Glu
405 410 415

Leu Arg Lys Lys Leu Ser Cys Lys Pro Phe Lys Trp Tyr Leu Glu Asn
420 425 430

Val Tyr Pro Glu Leu Arg Val Pro Asp His Gln Asp Ile Ala Phe Gly
435 440 445

Ala Leu Gln Gln Gly Thr Asn Cys Leu Asp Thr Leu Gly His Phe Ala
450 455 460

Asp Gly Val Val Gly Val Tyr Glu Cys His Asn Ala Gly Gly Asn Gln
465 470 475 480

Glu Trp Ala Leu Thr Lys Glu Lys Ser Val Lys His Met Asp Leu Cys

485 490 495

Leu Thr Val Val Asp Arg Ala Pro Gly Ser Leu Ile Lys Leu Gln Gly
500 505 510

Cys Arg Glu Asn Asp Ser Arg Gln Lys Trp Glu Gln Ile Glu Gly Asn
515 520 525

Ser Lys Leu Arg His Val Gly Ser Asn Leu Cys Leu Asp Ser Arg Thr
530 535 540

Ala Lys Ser Gly Gly Leu Ser Val Glu Val Cys Gly Pro Ala Leu Ser
545 550 555 560

Gln Gln Trp Lys Phe Thr Leu Asn Leu Gln Gln
565 570

<210> 3
<211> 1593
<212> DNA
<213> human
<220>
<221> misc_feature
<223> delat 40 GalNAcT2

<400> 3
aggaaggagg actggaatga aattgacccc attaaaaaga aagaccttca tcacagcaat 60

ggagaagaga aagcacaaag catggagacc ctccctccag ggaaagtacg gtggccagac 120

tttaaccagg aagcttatgt tggagggacg atggtccgct ccgggcagga cccttacgcc 180

cgcaacaagt tcaaccaggt ggagagtgat aagcttcgaa tggacagagc catccctgac 240

acccggcatg accagtgtca gcggaagcag tggcgggtgg atctgccggc caccagcgtg 300

gtgatcacgt ttcacaatga agccaggtcg gccctactca ggaccgtggt cagcgtgctt 360

aagaaaagcc cgccccatct cataaaagaa atcatcttgg tggatgacta cagcaatgat 420

cctgaggacg gggctctctt ggggaaaatt gagaaagtgc gagttcttag aaatgatcga 480

cgagaaggcc tcatgcgctc acgggttcgg ggggccgatg ctgcccaagc caaggtcctg 540

accttcctgg acagtcactg cgagtgtaat gagcactggc tggagcccct cctggaaagg 600

gtggcggagg acaggactcg ggttgtgtca cccatcatcg atgtcattaa tatggacaac 660

tttcagtatg tgggggcatc tgctgacttg aagggcggtt ttgattggaa cttggtattc 720

aagtgggatt acatgacgcc tgagcagaga aggtcccggc aggggaaccc agtcgccct 780

ataaaaaccc ccatgattgc tggtgggctg tttgtgatgg ataagttcta ttttgaagaa 840

ctggggaagt acgacatgat gatggatgtg tggggaggag agaacctaga gatctcgttc 900

cgcgtgtggc agtgtggtgg cagcctggag atcatcccgt gcagccgtgt gggacacgtg 960

ttccggaagc agcacccta cacgttcccg ggtggcagtg gcactgtctt tgcccgaaac 1020

acccgccggg cagcagaggt ctggatggat gaatacaaaa atttctatta tgcagcagtg 1080

ccttctgcta gaaacgttcc ttatggaaat attcagagca gattggagct taggaagaaa 1140

ctcagctgca agcctttcaa atggtacctt gaaaatgtct atccagagtt aagggttcca 1200

gaccatcagg atatagcttt tggggccttg cagcagggaa ctaactgcct cgacactttg 1260

ggacactttg ctgatggtgt ggttggagtt tatgaatgtc acaatgctgg gggaaaccag 1320

gaatgggcct tgacgaagga gaagtcggtg aagcacatgg atttgtgcct tactgtggtg 1380

gaccgggcac cgggctctct tataaagctg cagggctgcc gagaaaatga cagcagacag 1440

aaatgggaac agatcgaggg caactccaag ctgaggcacg tgggcagcaa cctgtgcctg 1500

gacagtcgca cggccaagag cgggggccta agcgtggagg tgtgtggccc ggcctttcg 1560

cagcagtgga agttcacgct caacctgcag cag 1593

<210> 4
<211> 531
<212> PRT
<213> human
<220>
<221> MISC_FEATURE
<223> delta 40 GalNAcT2

<400> 4

Arg Lys Glu Asp Trp Asn Glu Ile Asp Pro Ile Lys Lys Lys Asp Leu
1 5 10 15

His His Ser Asn Gly Glu Glu Lys Ala Gln Ser Met Glu Thr Leu Pro
20 25 30

Pro Gly Lys Val Arg Trp Pro Asp Phe Asn Gln Glu Ala Tyr Val Gly
35 40 45

Gly Thr Met Val Arg Ser Gly Gln Asp Pro Tyr Ala Arg Asn Lys Phe
50 55 60

Asn Gln Val Glu Ser Asp Lys Leu Arg Met Asp Arg Ala Ile Pro Asp
65 70 75 80

Thr Arg His Asp Gln Cys Gln Arg Lys Gln Trp Arg Val Asp Leu Pro
85 90 95

Ala Thr Ser Val Val Ile Thr Phe His Asn Glu Ala Arg Ser Ala Leu
100 105 110

Leu Arg Thr Val Val Ser Val Leu Lys Lys Ser Pro Pro His Leu Ile
115 120 125

Lys Glu Ile Ile Leu Val Asp Asp Tyr Ser Asn Asp Pro Glu Asp Gly
130 135 140

Ala Leu Leu Gly Lys Ile Glu Lys Val Arg Val Leu Arg Asn Asp Arg
145 150 155 160

Arg Glu Gly Leu Met Arg Ser Arg Val Arg Gly Ala Asp Ala Ala Gln
165 170 175

Ala Lys Val Leu Thr Phe Leu Asp Ser His Cys Glu Cys Asn Glu His
180 185 190

Trp Leu Glu Pro Leu Leu Glu Arg Val Ala Glu Asp Arg Thr Arg Val
195 200 205

Val Ser Pro Ile Ile Asp Val Ile Asn Met Asp Asn Phe Gln Tyr Val
210 215 220

Gly Ala Ser Ala Asp Leu Lys Gly Gly Phe Asp Trp Asn Leu Val Phe
225 230 235 240

Lys Trp Asp Tyr Met Thr Pro Glu Gln Arg Arg Ser Arg Gln Gly Asn
245 250 255

Pro Val Ala Pro Ile Lys Thr Pro Met Ile Ala Gly Gly Leu Phe Val
260 265 270

Met Asp Lys Phe Tyr Phe Glu Glu Leu Gly Lys Tyr Asp Met Met Met
275 280 285

Asp Val Trp Gly Gly Glu Asn Leu Glu Ile Ser Phe Arg Val Trp Gln
290 295 300

Cys Gly Gly Ser Leu Glu Ile Ile Pro Cys Ser Arg Val Gly His Val
305 310 315 320

Phe Arg Lys Gln His Pro Tyr Thr Phe Pro Gly Gly Ser Gly Thr Val
325 330 335

Phe Ala Arg Asn Thr Arg Arg Ala Ala Glu Val Trp Met Asp Glu Tyr
340 345 350

Lys Asn Phe Tyr Tyr Ala Ala Val Pro Ser Ala Arg Asn Val Pro Tyr
355 360 365

Gly Asn Ile Gln Ser Arg Leu Glu Leu Arg Lys Lys Leu Ser Cys Lys
370 375 380

Pro Phe Lys Trp Tyr Leu Glu Asn Val Tyr Pro Glu Leu Arg Val Pro
385 390 395 400

Asp His Gln Asp Ile Ala Phe Gly Ala Leu Gln Gln Gly Thr Asn Cys
405 410 415

Leu Asp Thr Leu Gly His Phe Ala Asp Gly Val Val Gly Val Tyr Glu
420 425 430

Cys His Asn Ala Gly Gly Asn Gln Glu Trp Ala Leu Thr Lys Glu Lys
435 440 445

Ser Val Lys His Met Asp Leu Cys Leu Thr Val Val Asp Arg Ala Pro
450 455 460

Gly Ser Leu Ile Lys Leu Gln Gly Cys Arg Glu Asn Asp Ser Arg Gln
465 470 475 480

Lys Trp Glu Gln Ile Glu Gly Asn Ser Lys Leu Arg His Val Gly Ser

485 490 495

Asn Leu Cys Leu Asp Ser Arg Thr Ala Lys Ser Gly Gly Leu Ser Val
500 505 510

Glu Val Cys Gly Pro Ala Leu Ser Gln Gln Trp Lys Phe Thr Leu Asn
515 520 525

Leu Gln Gln
530

<210> 5
<211> 1560
<212> DNA
<213> human
<220>
<221> misc_feature
<223> delta 51 GalNAcT2

<400> 5
aaaaagaaag accttcatca cagcaatgga gaagagaaag cacaaagcat ggagaccctc 60

cctccaggga aagtacggtg gccagacttt aaccaggaag cttatgttgg agggacgatg 120

gtccgctccg ggcaggaccc ttacgcccgc aacaagttca accaggtgga gagtgataag 180

cttcgaatgg acagagccat ccctgacacc cggcatgacc agtgtcagcg gaagcagtgg 240

cgggtggatc tgccggccac cagcgtggtg atcacgtttc acaatgaagc caggtcggcc 300

ctactcagga ccgtggtcag cgtgcttaag aaaagcccgc cccatctcat aaaagaaatc 360

atcttggtgg atgactacag caatgatcct gaggacgggg ctctcttggg gaaaattgag 420

aaagtgcgag ttcttagaaa tgatcgacga gaaggcctca tgcgctcacg ggttcggggg 480

gccgatgctg cccaagccaa ggtcctgacc ttcctggaca gtcactgcga gtgtaatgag 540

cactggctgg agcccctcct ggaaagggtg gcggaggaca ggactcgggt tgtgtcaccc 600

atcatcgatg tcattaatat ggacaacttt cagtatgtgg gggcatctgc tgacttgaag 660

ggcggttttg attggaactt ggtattcaag tgggattaca tgacgcctga gcagagaagg 720

tcccggcagg ggaacccagt cgcccctata aaaaccccca tgattgctgg tgggctgttt 780

gtgatggata agttctattt tgaagaactg gggaagtacg acatgatgat ggatgtgtgg 840

ggaggagaga acctagagat ctcgttccgc gtgtggcagt gtggtggcag cctggagatc 900

atcccgtgca gccgtgtggg acacgtgttc cggaagcagc acccctacac gttcccgggt 960

ggcagtggca ctgtctttgc ccgaaacacc cgccgggcag cagaggtctg gatggatgaa 1020

tacaaaaatt tctattatgc agcagtgcct tctgctagaa acgttcctta tggaaatatt 1080

cagagcagat tggagcttag gaagaaactc agctgcaagc ctttcaaatg gtaccttgaa 1140

aatgtctatc cagagttaag ggttccagac catcaggata tagcttttgg ggccttgcag 1200

cagggaacta actgcctcga cactttggga cactttgctg atggtgtggt tggagtttat 1260

gaatgtcaca atgctggggg aaaccaggaa tgggccttga cgaaggagaa gtcggtgaag 1320

cacatggatt tgtgccttac tgtggtggac cgggcaccgg gctctcttat aaagctgcag 1380

ggctgccgag aaaatgacag cagacagaaa tgggaacaga tcgagggcaa ctccaagctg 1440

aggcacgtgg gcagcaacct gtgcctggac agtcgcacgg ccaagagcgg gggcctaagc 1500

gtggaggtgt gtggcccggc cctttcgcag cagtggaagt tcacgctcaa cctgcagcag 1560

<210> 6
<211> 520
<212> PRT
<213> human
<220>
<221> MISC_FEATURE
<223> delta 51 GalNAcT2

<400> 6

Lys Lys Lys Asp Leu His His Ser Asn Gly Glu Glu Lys Ala Gln Ser
1 5 10 15

Met Glu Thr Leu Pro Pro Gly Lys Val Arg Trp Pro Asp Phe Asn Gln
20 25 30

Glu Ala Tyr Val Gly Gly Thr Met Val Arg Ser Gly Gln Asp Pro Tyr
35 40 45

Ala Arg Asn Lys Phe Asn Gln Val Glu Ser Asp Lys Leu Arg Met Asp
50 55 60

Arg Ala Ile Pro Asp Thr Arg His Asp Gln Cys Gln Arg Lys Gln Trp
65 70 75 80

Arg Val Asp Leu Pro Ala Thr Ser Val Val Ile Thr Phe His Asn Glu
85 90 95

Ala Arg Ser Ala Leu Leu Arg Thr Val Val Ser Val Leu Lys Lys Ser
100 105 110

Pro Pro His Leu Ile Lys Glu Ile Ile Leu Val Asp Asp Tyr Ser Asn
115 120 125

Asp Pro Glu Asp Gly Ala Leu Leu Gly Lys Ile Glu Lys Val Arg Val
130 135 140

Leu Arg Asn Asp Arg Arg Glu Gly Leu Met Arg Ser Arg Val Arg Gly
145 150 155 160

Ala Asp Ala Ala Gln Ala Lys Val Leu Thr Phe Leu Asp Ser His Cys
165 170 175

Glu Cys Asn Glu His Trp Leu Glu Pro Leu Leu Glu Arg Val Ala Glu
180 185 190

Asp Arg Thr Arg Val Val Ser Pro Ile Ile Asp Val Ile Asn Met Asp

```
Asn Phe Gln Tyr Val Gly Ala Ser Ala Asp Leu Lys Gly Gly Phe Asp
    210                 215                 220

Trp Asn Leu Val Phe Lys Trp Asp Tyr Met Thr Pro Glu Gln Arg Arg
225                 230                 235                 240

Ser Arg Gln Gly Asn Pro Val Ala Pro Ile Lys Thr Pro Met Ile Ala
                245                 250                 255

Gly Gly Leu Phe Val Met Asp Lys Phe Tyr Phe Glu Glu Leu Gly Lys
            260                 265                 270

Tyr Asp Met Met Met Asp Val Trp Gly Gly Glu Asn Leu Glu Ile Ser
        275                 280                 285

Phe Arg Val Trp Gln Cys Gly Gly Ser Leu Glu Ile Ile Pro Cys Ser
    290                 295                 300

Arg Val Gly His Val Phe Arg Lys Gln His Pro Tyr Thr Phe Pro Gly
305                 310                 315                 320

Gly Ser Gly Thr Val Phe Ala Arg Asn Thr Arg Arg Ala Ala Glu Val
                325                 330                 335

Trp Met Asp Glu Tyr Lys Asn Phe Tyr Tyr Ala Ala Val Pro Ser Ala
            340                 345                 350

Arg Asn Val Pro Tyr Gly Asn Ile Gln Ser Arg Leu Glu Leu Arg Lys
        355                 360                 365

Lys Leu Ser Cys Lys Pro Phe Lys Trp Tyr Leu Glu Asn Val Tyr Pro
    370                 375                 380

Glu Leu Arg Val Pro Asp His Gln Asp Ile Ala Phe Gly Ala Leu Gln
385                 390                 395                 400

Gln Gly Thr Asn Cys Leu Asp Thr Leu Gly His Phe Ala Asp Gly Val
                405                 410                 415

Val Gly Val Tyr Glu Cys His Asn Ala Gly Gly Asn Gln Glu Trp Ala
            420                 425                 430

Leu Thr Lys Glu Lys Ser Val Lys His Met Asp Leu Cys Leu Thr Val
        435                 440                 445

Val Asp Arg Ala Pro Gly Ser Leu Ile Lys Leu Gln Gly Cys Arg Glu
    450                 455                 460

Asn Asp Ser Arg Gln Lys Trp Glu Gln Ile Glu Gly Asn Ser Lys Leu
465                 470                 475                 480

Arg His Val Gly Ser Asn Leu Cys Leu Asp Ser Arg Thr Ala Lys Ser
                485                 490                 495

Gly Gly Leu Ser Val Glu Val Cys Gly Pro Ala Leu Ser Gln Gln Trp
            500                 505                 510

Lys Phe Thr Leu Asn Leu Gln Gln
        515                 520
```

<210> 7
<211> 1494
<212> DNA
<213> human
<220>
<221> misc_feature
<223> delta 73 GalNAcT2

<400> 7
gggaaagtac ggtggccaga ctttaaccag gaagcttatg ttggagggac gatggtccgc 60

tccgggcagg acccttacgc ccgcaacaag ttcaaccagg tggagagtga taagcttcga 120

atggacagag ccatccctga cacccggcat gaccagtgtc agcggaagca gtggcgggtg 180

gatctgccgg ccaccagcgt ggtgatcacg tttcacaatg aagccaggtc ggccctactc 240

aggaccgtgg tcagcgtgct taagaaaagc ccgccccatc tcataaaaga aatcatcttg 300

gtggatgact acagcaatga tcctgaggac ggggctctct tggggaaaat tgagaaagtg 360

cgagttctta gaaatgatcg acgagaaggc ctcatgcgct cacgggttcg gggggccgat 420

gctgcccaag ccaaggtcct gaccttcctg gacagtcact gcgagtgtaa tgagcactgg 480

ctggagcccc tcctggaaag ggtggcggag gacaggactc gggttgtgtc acccatcatc 540

gatgtcatta atatggacaa ctttcagtat gtgggggcat ctgctgactt gaagggcggt 600

tttgattgga acttggtatt caagtgggat tacatgacgc ctgagcagag aaggtcccgg 660

caggggaacc cagtcgcccc tataaaaacc cccatgattg ctggtgggct gtttgtgatg 720

gataagttct attttgaaga actggggaag tacgacatga tgatggatgt gtggggagga 780

gagaacctag agatctcgtt ccgcgtgtgg cagtgtggtg gcagcctgga gatcatcccg 840

tgcagccgtg tgggacacgt gttccggaag cagcacccct acacgttccc gggtggcagt 900

ggcactgtct ttgcccgaaa cacccgccgg gcagcagagg tctggatgga tgaatacaaa 960

aatttctatt atgcagcagt gccttctgct agaaacgttc cttatggaaa tattcagagc 1020

agattggagc ttaggaagaa actcagctgc aagcctttca aatggtacct tgaaaatgtc 1080

tatccagagt taagggttcc agaccatcag gatatagctt ttggggcctt gcagcaggga 1140

actaactgcc tcgacacttt gggacacttt gctgatggtg tggttggagt ttatgaatgt 1200

cacaatgctg ggggaaacca ggaatgggcc ttgacgaagg agaagtcggt gaagcacatg 1260

gatttgtgcc ttactgtggt ggaccgggca ccgggctctc ttataaagct gcagggctgc 1320

cgagaaaatg acagcagaca gaaatgggaa cagatcgagg gcaactccaa gctgaggcac 1380

gtgggcagca acctgtgcct ggacagtcgc acggccaaga gcgggggcct aagcgtggag 1440

gtgtgtggcc cggccctttc gcagcagtgg aagttcacgc tcaacctgca gcag 1494

<210> 8
<211> 498
<212> PRT
<213> human
<220>
<221> MISC_FEATURE
<223> delta 73 GalNAcT2

<400> 8

Gly Lys Val Arg Trp Pro Asp Phe Asn Gln Glu Ala Tyr Val Gly Gly
1 5 10 15

Thr Met Val Arg Ser Gly Gln Asp Pro Tyr Ala Arg Asn Lys Phe Asn
20 25 30

Gln Val Glu Ser Asp Lys Leu Arg Met Asp Arg Ala Ile Pro Asp Thr
35 40 45

Arg His Asp Gln Cys Gln Arg Lys Gln Trp Arg Val Asp Leu Pro Ala
50 55 60

Thr Ser Val Val Ile Thr Phe His Asn Glu Ala Arg Ser Ala Leu Leu
65 70 75 80

Arg Thr Val Val Ser Val Leu Lys Lys Ser Pro Pro His Leu Ile Lys
85 90 95

Glu Ile Ile Leu Val Asp Asp Tyr Ser Asn Asp Pro Glu Asp Gly Ala
100 105 110

Leu Leu Gly Lys Ile Glu Lys Val Arg Val Leu Arg Asn Asp Arg Arg
115 120 125

Glu Gly Leu Met Arg Ser Arg Val Arg Gly Ala Asp Ala Ala Gln Ala
130 135 140

Lys Val Leu Thr Phe Leu Asp Ser His Cys Glu Cys Asn Glu His Trp
145 150 155 160

Leu Glu Pro Leu Leu Glu Arg Val Ala Glu Asp Arg Thr Arg Val Val
165 170 175

Ser Pro Ile Ile Asp Val Ile Asn Met Asp Asn Phe Gln Tyr Val Gly
180 185 190

Ala Ser Ala Asp Leu Lys Gly Gly Phe Asp Trp Asn Leu Val Phe Lys
195 200 205

Trp Asp Tyr Met Thr Pro Glu Gln Arg Arg Ser Arg Gln Gly Asn Pro
210 215 220

Val Ala Pro Ile Lys Thr Pro Met Ile Ala Gly Gly Leu Phe Val Met
225 230 235 240

Asp Lys Phe Tyr Phe Glu Glu Leu Gly Lys Tyr Asp Met Met Met Asp
245 250 255

Val Trp Gly Gly Glu Asn Leu Glu Ile Ser Phe Arg Val Trp Gln Cys
260 265 270

Gly Gly Ser Leu Glu Ile Ile Pro Cys Ser Arg Val Gly His Val Phe
275 280 285

Arg Lys Gln His Pro Tyr Thr Phe Pro Gly Gly Ser Gly Thr Val Phe
290 295 300

Ala Arg Asn Thr Arg Arg Ala Ala Glu Val Trp Met Asp Glu Tyr Lys
305 310 315 320

Asn Phe Tyr Tyr Ala Ala Val Pro Ser Ala Arg Asn Val Pro Tyr Gly
325 330 335

Asn Ile Gln Ser Arg Leu Glu Leu Arg Lys Lys Leu Ser Cys Lys Pro
340 345 350

Phe Lys Trp Tyr Leu Glu Asn Val Tyr Pro Glu Leu Arg Val Pro Asp
355 360 365

His Gln Asp Ile Ala Phe Gly Ala Leu Gln Gln Gly Thr Asn Cys Leu
370 375 380

Asp Thr Leu Gly His Phe Ala Asp Gly Val Val Gly Val Tyr Glu Cys
385 390 395 400

His Asn Ala Gly Gly Asn Gln Glu Trp Ala Leu Thr Lys Glu Lys Ser
405 410 415

Val Lys His Met Asp Leu Cys Leu Thr Val Val Asp Arg Ala Pro Gly
420 425 430

Ser Leu Ile Lys Leu Gln Gly Cys Arg Glu Asn Asp Ser Arg Gln Lys
435 440 445

Trp Glu Gln Ile Glu Gly Asn Ser Lys Leu Arg His Val Gly Ser Asn
450 455 460

Leu Cys Leu Asp Ser Arg Thr Ala Lys Ser Gly Gly Leu Ser Val Glu
465 470 475 480

Val Cys Gly Pro Ala Leu Ser Gln Gln Trp Lys Phe Thr Leu Asn Leu
485 490 495

Gln Gln

<210> 9
<211> 1431
<212> DNA
<213> human
<220>
<221> misc_feature
<223> delta 94 GalNAcT2

<400> 9
gggcaggacc cttacgcccg caacaagttc aaccaggtgg agagtgataa gcttcgaatg 60

gacagagcca tccctgacac ccggcatgac cagtgtcagc ggaagcagtg gcgggtggat 120

ctgccggcca ccagcgtggt gatcacgttt cacaatgaag ccaggtcggc cctactcagg 180

accgtggtca gcgtgcttaa gaaaagcccg ccccatctca taaaagaaat catcttggtg 240

gatgactaca gcaatgatcc tgaggacggg gctctcttgg ggaaaattga gaaagtgcga 300

gttcttagaa atgatcgacg agaaggcctc atgcgctcac gggttcgggg ggccgatgct 360

gcccaagcca aggtcctgac cttcctggac agtcactgcg agtgtaatga gcactggctg 420

gagcccctcc tggaaagggt ggcggaggac aggactcggg ttgtgtcacc catcatcgat 480

gtcattaata tggacaactt tcagtatgtg ggggcatctg ctgacttgaa gggcggtttt 540

gattggaact tggtattcaa gtgggattac atgacgcctg agcagagaag gtcccggcag 600

gggaacccag tcgcccctat aaaaaccccc atgattgctg gtgggctgtt tgtgatggat 660

aagttctatt ttgaagaact ggggaagtac gacatgatga tggatgtgtg gggaggagag 720

aacctagaga tctcgttccg cgtgtggcag tgtggtggca gcctggagat catcccgtgc 780

agccgtgtgg gacacgtgtt ccggaagcag cacccctaca cgttcccggg tggcagtggc 840

actgtctttg cccgaaacac ccgccgggca gcagaggtct ggatggatga atacaaaaat 900

ttctattatg cagcagtgcc ttctgctaga aacgttcctt atggaaatat tcagagcaga 960

ttggagctta ggaagaaact cagctgcaag cctttcaaat ggtaccttga aaatgtctat 1020

ccagagttaa gggttccaga ccatcaggat atagcttttg gggccttgca gcagggaact 1080

aactgcctcg acactttggg acactttgct gatggtgtgg ttggagttta tgaatgtcac 1140

aatgctgggg gaaaccagga atgggccttg acgaaggaga agtcggtgaa gcacatggat 1200

ttgtgcctta ctgtggtgga ccgggcaccg ggctctctta taaagctgca gggctgccga 1260

gaaaatgaca gcagacagaa atgggaacag atcgagggca actccaagct gaggcacgtg 1320

ggcagcaacc tgtgcctgga cagtcgcacg gccaagagcg ggggcctaag cgtggaggtg 1380

tgtggcccgg ccctttcgca gcagtggaag ttcacgctca acctgcagca g 1431

<210> 10
<211> 477
<212> PRT
<213> human
<220>
<221> MISC_FEATURE
<223> delta 94 GalNAcT2

<400> 10

Gly Gln Asp Pro Tyr Ala Arg Asn Lys Phe Asn Gln Val Glu Ser Asp
1 5 10 15

Lys Leu Arg Met Asp Arg Ala Ile Pro Asp Thr Arg His Asp Gln Cys
20 25 30

Gln Arg Lys Gln Trp Arg Val Asp Leu Pro Ala Thr Ser Val Val Ile
35 40 45

Thr Phe His Asn Glu Ala Arg Ser Ala Leu Leu Arg Thr Val Val Ser
50 55 60

Val Leu Lys Lys Ser Pro Pro His Leu Ile Lys Glu Ile Ile Leu Val
65 70 75 80

Asp Asp Tyr Ser Asn Asp Pro Glu Asp Gly Ala Leu Leu Gly Lys Ile
85 90 95

Glu Lys Val Arg Val Leu Arg Asn Asp Arg Arg Glu Gly Leu Met Arg
100 105 110

Ser Arg Val Arg Gly Ala Asp Ala Ala Gln Ala Lys Val Leu Thr Phe
115 120 125

Leu Asp Ser His Cys Glu Cys Asn Glu His Trp Leu Glu Pro Leu Leu
130 135 140

Glu Arg Val Ala Glu Asp Arg Thr Arg Val Val Ser Pro Ile Ile Asp
145 150 155 160

Val Ile Asn Met Asp Asn Phe Gln Tyr Val Gly Ala Ser Ala Asp Leu
165 170 175

Lys Gly Gly Phe Asp Trp Asn Leu Val Phe Lys Trp Asp Tyr Met Thr
180 185 190

Pro Glu Gln Arg Arg Ser Arg Gln Gly Asn Pro Val Ala Pro Ile Lys
195 200 205

Thr Pro Met Ile Ala Gly Gly Leu Phe Val Met Asp Lys Phe Tyr Phe
210 215 220

Glu Glu Leu Gly Lys Tyr Asp Met Met Met Asp Val Trp Gly Gly Glu
225 230 235 240

Asn Leu Glu Ile Ser Phe Arg Val Trp Gln Cys Gly Gly Ser Leu Glu
245 250 255

Ile Ile Pro Cys Ser Arg Val Gly His Val Phe Arg Lys Gln His Pro
260 265 270

Tyr Thr Phe Pro Gly Gly Ser Gly Thr Val Phe Ala Arg Asn Thr Arg
275 280 285

Arg Ala Ala Glu Val Trp Met Asp Glu Tyr Lys Asn Phe Tyr Tyr Ala
290 295 300

Ala Val Pro Ser Ala Arg Asn Val Pro Tyr Gly Asn Ile Gln Ser Arg
305 310 315 320

Leu Glu Leu Arg Lys Lys Leu Ser Cys Lys Pro Phe Lys Trp Tyr Leu
325 330 335

Glu Asn Val Tyr Pro Glu Leu Arg Val Pro Asp His Gln Asp Ile Ala
340 345 350

Phe Gly Ala Leu Gln Gln Gly Thr Asn Cys Leu Asp Thr Leu Gly His
355 360 365

Phe Ala Asp Gly Val Val Gly Val Tyr Glu Cys His Asn Ala Gly Gly
370 375 380

Asn Gln Glu Trp Ala Leu Thr Lys Glu Lys Ser Val Lys His Met Asp
385 390 395 400

Leu Cys Leu Thr Val Val Asp Arg Ala Pro Gly Ser Leu Ile Lys Leu
405 410 415

Gln Gly Cys Arg Glu Asn Asp Ser Arg Gln Lys Trp Glu Gln Ile Glu
420 425 430

Gly Asn Ser Lys Leu Arg His Val Gly Ser Asn Leu Cys Leu Asp Ser
435 440 445

Arg Thr Ala Lys Ser Gly Gly Leu Ser Val Glu Val Cys Gly Pro Ala
450 455 460

Leu Ser Gln Gln Trp Lys Phe Thr Leu Asn Leu Gln Gln
465 470 475

<210> 11
<211> 28
<212> DNA
<213> artificial sequence
<220>
<223> N41R primer

<400> 11
cgcggatcca ggaaggagga ctggaatg 28

<210> 12
<211> 33
<212> DNA
<213> artificial sequence
<220>
<223> N52K primer

<400> 12
cgcggatcca aaaagaaaga ccttcatcac agc 33

<210> 13
<211> 30
<212> DNA
<213> artificial sequence
<220>
<223> N74G primer

<400> 13
cgcggatccg ggaaagtacg gtggccagac 30

<210> 14
<211> 27
<212> DNA
<213> artificial sequence
<220>
<223> N95G primer

<400> 14
cgcggatccg ggcaggaccc ttacgcc 27

<210> 15
<211> 29
<212> DNA
<213> artificial sequence
<220>
<223> Antisense Primer with STOP codon

<400> 15
ctgctcgagc tactgctgca ggttgagcg 29

<210> 16
<211> 10
<212> PRT
<213> artificial sequence
<220>
<223> MuC-2 - like peptide

<400> 16

Met Val Thr Pro Thr Pro Thr Pro Thr Cys
1 5 10

WHAT IS CLAIMED IS:

1. An isolated nucleic acid comprising a nucleic acid sequence encoding a truncated human GalNAcT2 polypeptide, wherein said truncated human GalNAcT2 polypeptide is lacking all or a portion of the GalNAcT2 signal domain, with the proviso that the encoded polypeptide is not a human GalNAcT2 truncation mutant polypeptide lacking amino acid residues 1-51.

2. The isolated nucleic acid of claim 1, wherein said truncated human GalNAcT2 polypeptide is further lacking all or a portion the GalNAcT2 transmembrane domain, with the proviso that the encoded polypeptide is not a human GalNAcT2 truncation mutant polypeptide lacking amino acid residues 1-51.

3. The isolated nucleic of claim 2, wherein said truncated human GalNAcT2 polypeptide is further lacking all or a portion the GalNAcT2 stem domain, with the proviso that the encoded polypeptide is not a human GalNAcT2 truncation mutant polypeptide lacking amino acid residues 1-51.

4. The isolated nucleic acid of claim 1, comprising a nucleic acid sequence encoding a truncated human GalNAcT2 polypeptide, said nucleic acid sequence having at least 90% identity with a nucleic acid selected from the group consisting of SEQ ID NO:3, SEQ ID NO:7 and SEQ ID NO:9

5. The isolated nucleic acid of claim 4, said isolated nucleic acid comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:7 and SEQ ID NO:9.

6. An isolated nucleic acid of claim 4, e said isolated nucleic acid consisting of a nucleic acid sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:7 and SEQ ID NO:9.

7. An isolated chimeric nucleic acid encoding a fusion polypeptide, said fusion polypeptide comprising a tag polypeptide covalently linked to a second polypeptide encoded by the isolated nucleic acid of claim 1.

8. The isolated chimeric nucleic acid of claim 7, wherein said tag polypeptide is selected from the group consisting of a maltose binding protein, a histidine tag,

a Factor IX tag, a glutathione-S-transferase tag, a FLAG-tag, and a starch binding domain tag.

9. An isolated truncated human GalNAcT2 polypeptide, wherein said truncated human GalNAcT2 polypeptide is lacking all or a portion of the GalNAcT2 signal domain, with the proviso that said polypeptide is not a human GalNAcT2 polypeptide truncation mutant lacking amino acid residues 1-51.

10. The isolated truncated human GalNAcT2 polypeptide of claim 9, wherein said truncated human GalNAcT2 polypeptide is further lacking all or a portion the GalNAcT2 transmembrane domain, with the proviso that said polypeptide is not a human GalNAcT2 polypeptide truncation mutant lacking amino acid residues 1-51.

11. The isolated truncated human GalNAcT2 polypeptide of claim 10, wherein said truncated human GalNAcT2 polypeptide is further lacking all or a portion the GalNAcT2 stem domain, with the proviso that said polypeptide is not a human GalNAcT2 polypeptide truncation mutant lacking amino acid residues 1-51.

12. The isolated truncated human GalNAcT2 polypeptide of claim 9, having at least 90% identity with a polypeptide selected from the group consisting of SEQ ID NO:4, SEQ ID NO:8 and SEQ ID NO:10.

13. The isolated truncated human GalNAcT2 polypeptide of claim 9, comprising an amino acid sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:8 and SEQ ID NO:10.

14. The isolated truncated human GalNAcT2 polypeptide of claim 9, consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:8 and SEQ ID NO:10.

15. An isolated chimeric polypeptide comprising a tag polypeptide covalently linked to the isolated truncated GalNAcT2 polypeptide of claim 9.

16. The isolated chimeric polypeptide of claim 15, wherein said tag polypeptide is selected from the group consisting of a maltose binding protein, a histidine tag, a Factor IX tag, a glutathione-S-transferase tag, a FLAG-tag, and a starch binding domain tag.

17. The isolated nucleic acid of any one of claim 1, said nucleic acid further comprising a promoter/regulatory sequence operably linked thereto.

18. An expression vector comprising the isolated nucleic acid of claim 1.

19. A recombinant cell comprising the isolated expression vector of claim 18.

20. A recombinant cell of claim 19, wherein said recombinant cell is a eukaryotic cell or a prokaryotic cell.

21. The recombinant cell of claim 20, wherein said eukaryotic cell is selected from the group consisting of a mammalian cell, an insect cell, and a fungal cell.

22. The recombinant cell of claim 21, wherein said insect cell is selected from the group consisting of an SF9 cell, an SF9+ cell, an Sf21 cell, a HIGH FIVE cell or Drosophila Schneider S2 cell.

23. The recombinant cell of claim 20, wherein said prokaryotic cell is selected from the group consisting of an E. coli cell and a B. subtilis cell.

24. A method of producing a truncated human GalNAcT2 polypeptide, the method comprising growing the recombinant cell of claim 20 under conditions suitable for expression of the truncated human GalNAcT2 polypeptide.

25. A method of catalyzing the transfer of a GalNAc moiety to an acceptor moiety comprising incubating the polypeptide of claim 9 with a GalNAc moiety and an acceptor moiety, wherein said polypeptide mediates the covalent linkage of said GalNAc moiety to said acceptor moiety, thereby catalyzing the transfer of a GalNAc moiety to an acceptor moiety to produce a product saccharide, or a product glycoprotein, or a product glycopeptide.

26. The method of claim 25, wherein said acceptor moiety is a granulocyte colony stimulating factor (G-CSF) protein.

27. The method of claim 25, wherein said acceptor moiety is selected from the group consisting of erythropoietin, human growth hormone, granulocyte colony

stimulating factor, interferons alpha, -beta, and –gamma, Factor IX, follicle stimulating hormone, interleukin-2, erythropoietin, anti-TNF-alpha, and a lysosomal hydrolase.

28. The method of claim 25, wherein said polypeptide acceptor is a glycopeptide.

29. The method of claim 25, further wherein said GalNAc moiety comprises a polyethylene glycol moiety.

30. The method of claim 25, wherein the product saccharide, product glycoprotein, or product glycopeptide is produced on a commercial scale.

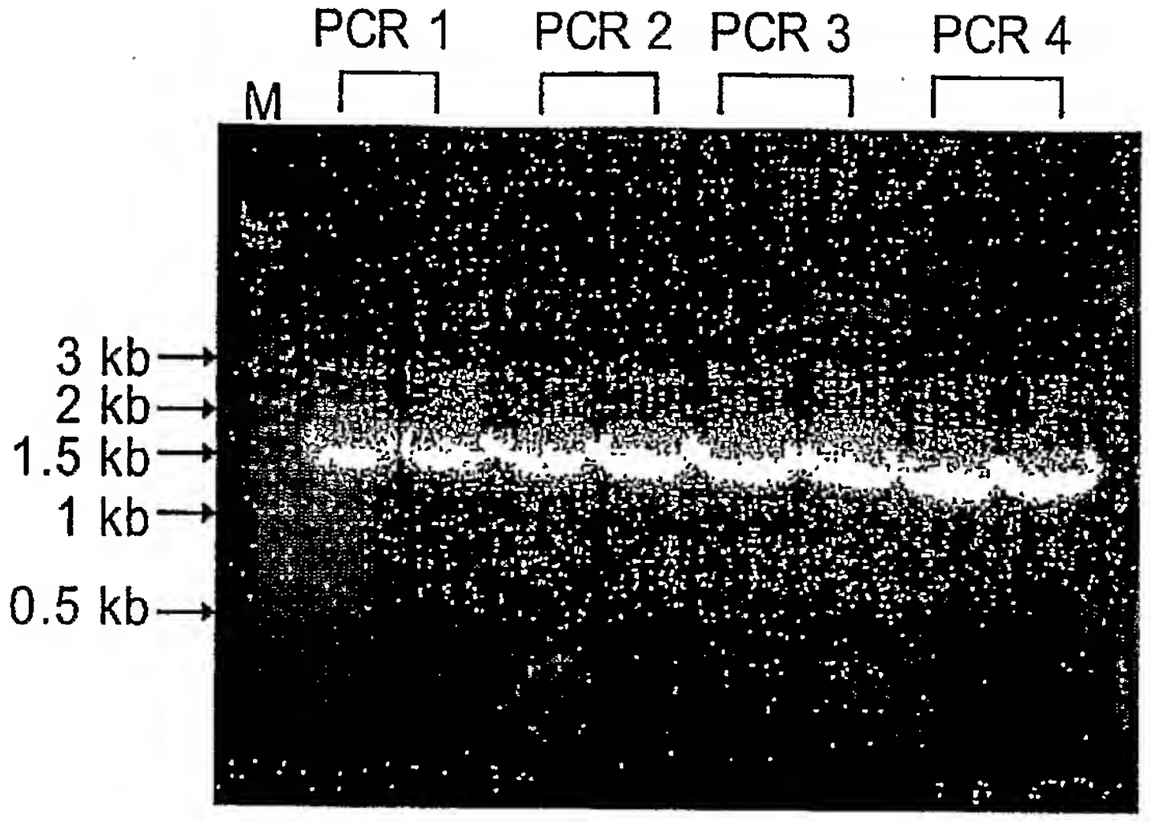


FIG. 1

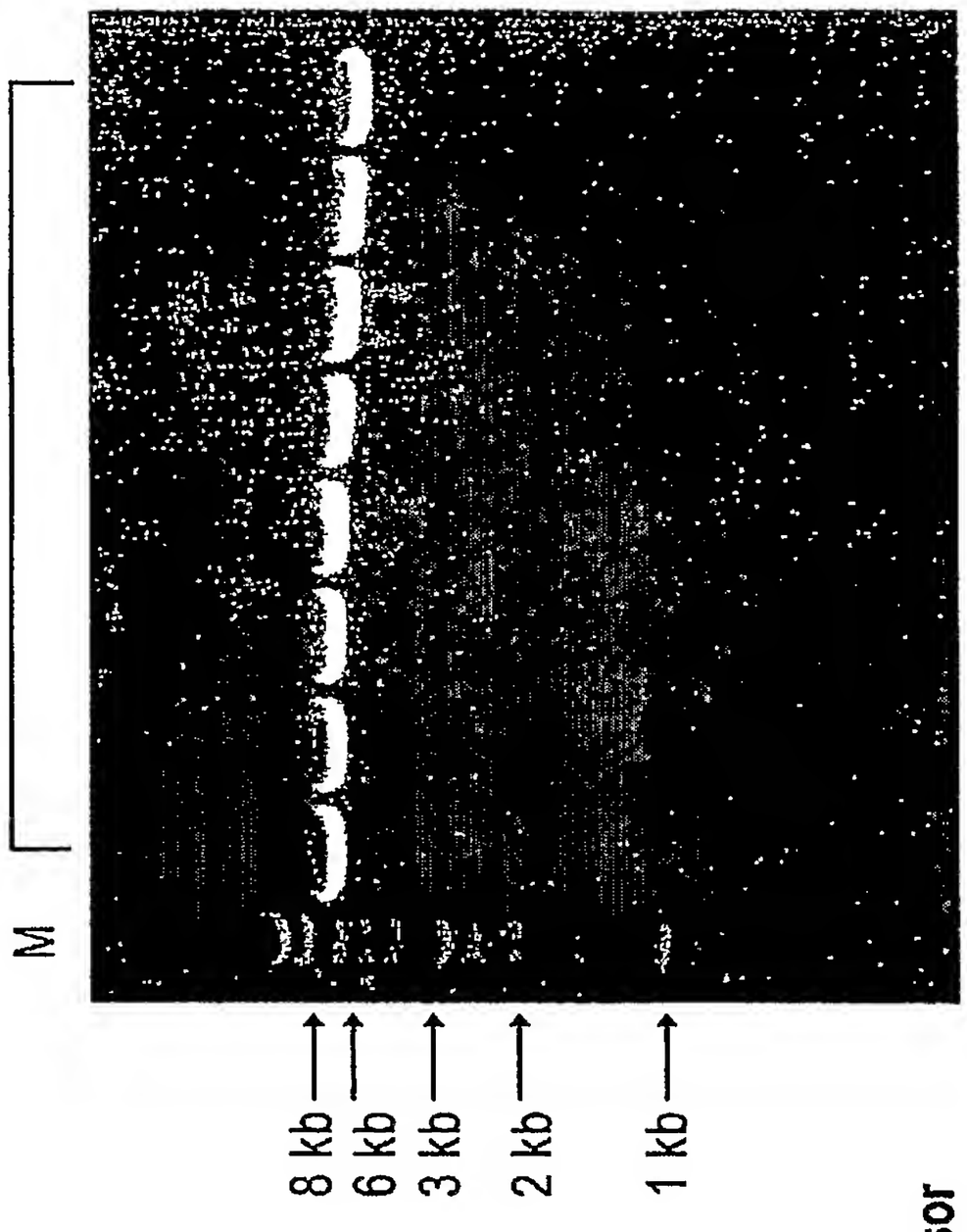

1
M
8 kb
6 kb
3 kb
2 kb
1 kb

FIG. 2B

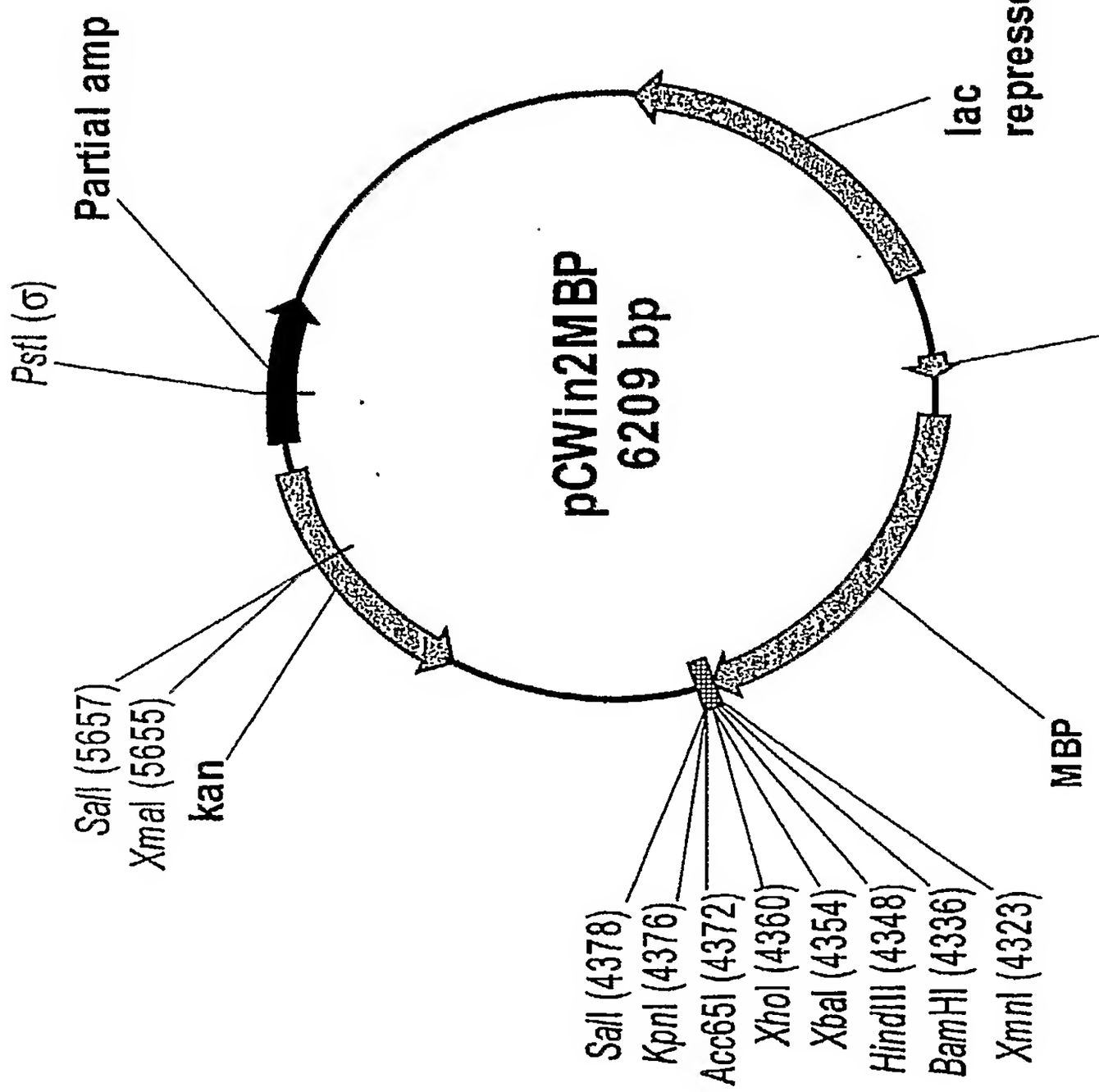

Partial amp
PstI (σ)
SalI (5657)
XmaI (5655)
kan
lac repressor
tac promoter
MBP
pCWin2MBP
6209 bp
SalI (4378)
KpnI (4376)
Acc65I (4372)
XhoI (4360)
XbaI (4354)
HindIII (4348)
BamHI (4336)
XmnI (4323)

FIG. 2A

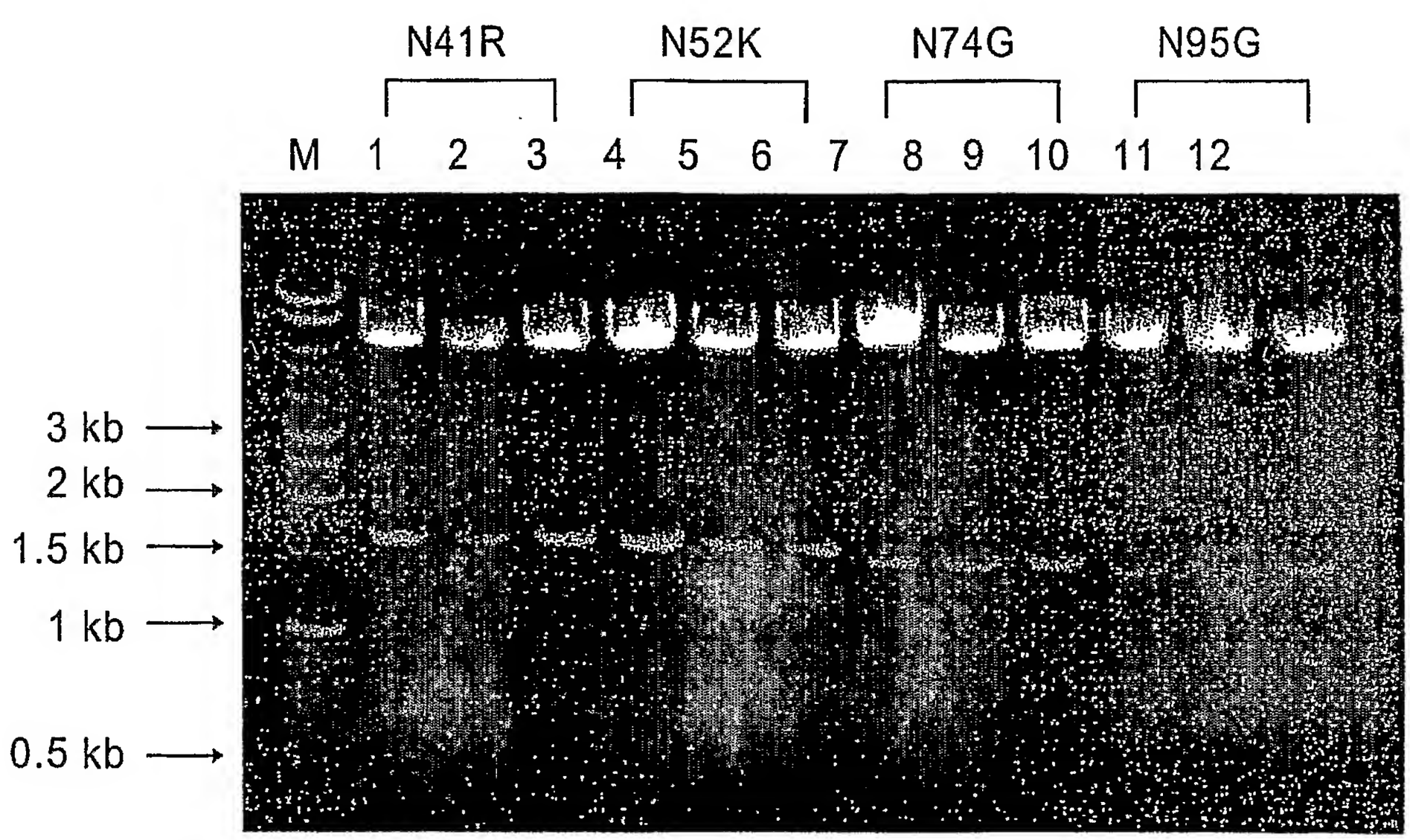


FIG. 3

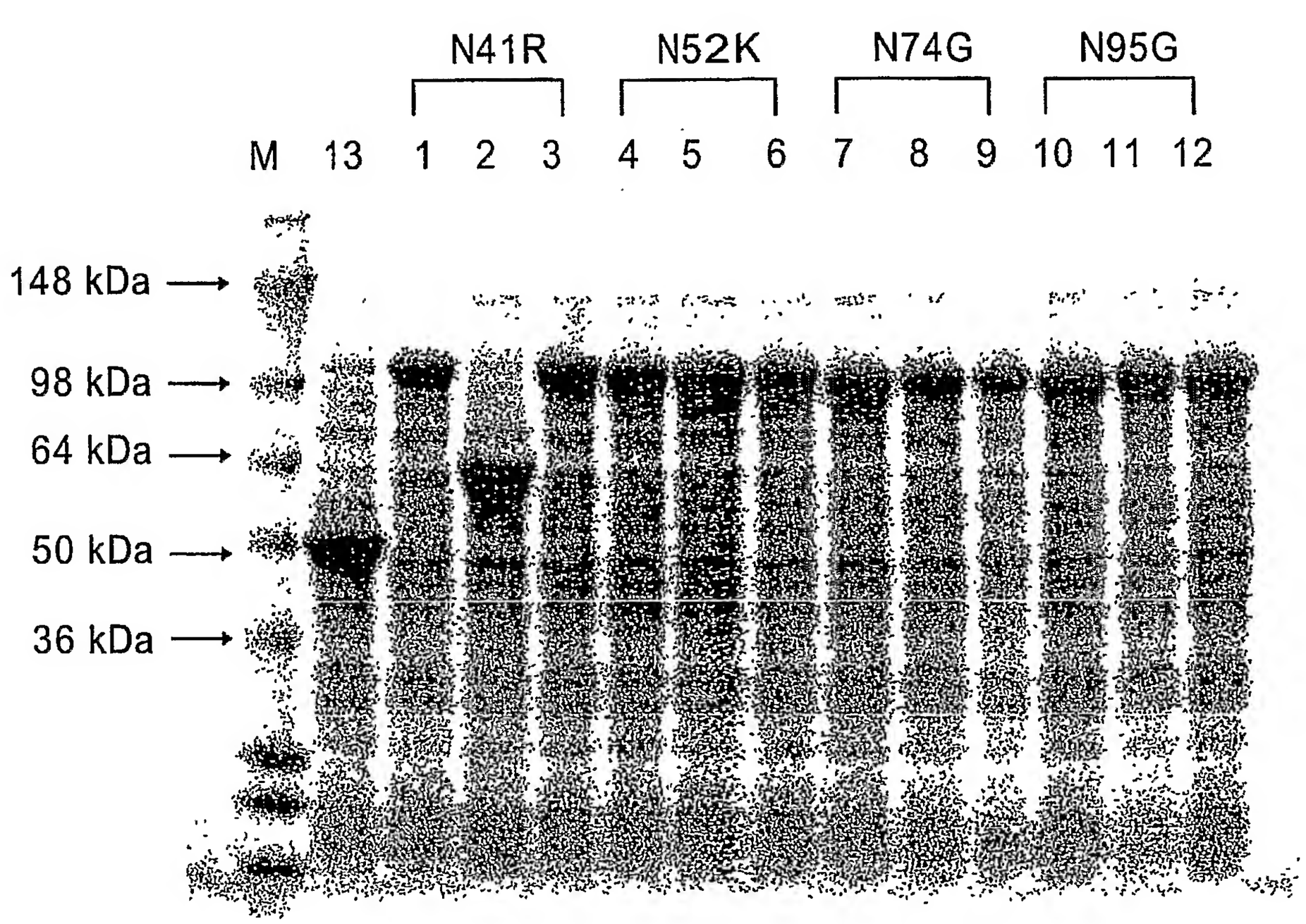

N41R
N52K
N74G
N95G
M 13 1 2 3 4 5 6 7 8 9 10 11 12
148 kDa
98 kDa
64 kDa
50 kDa
36 kDa

FIG. 4

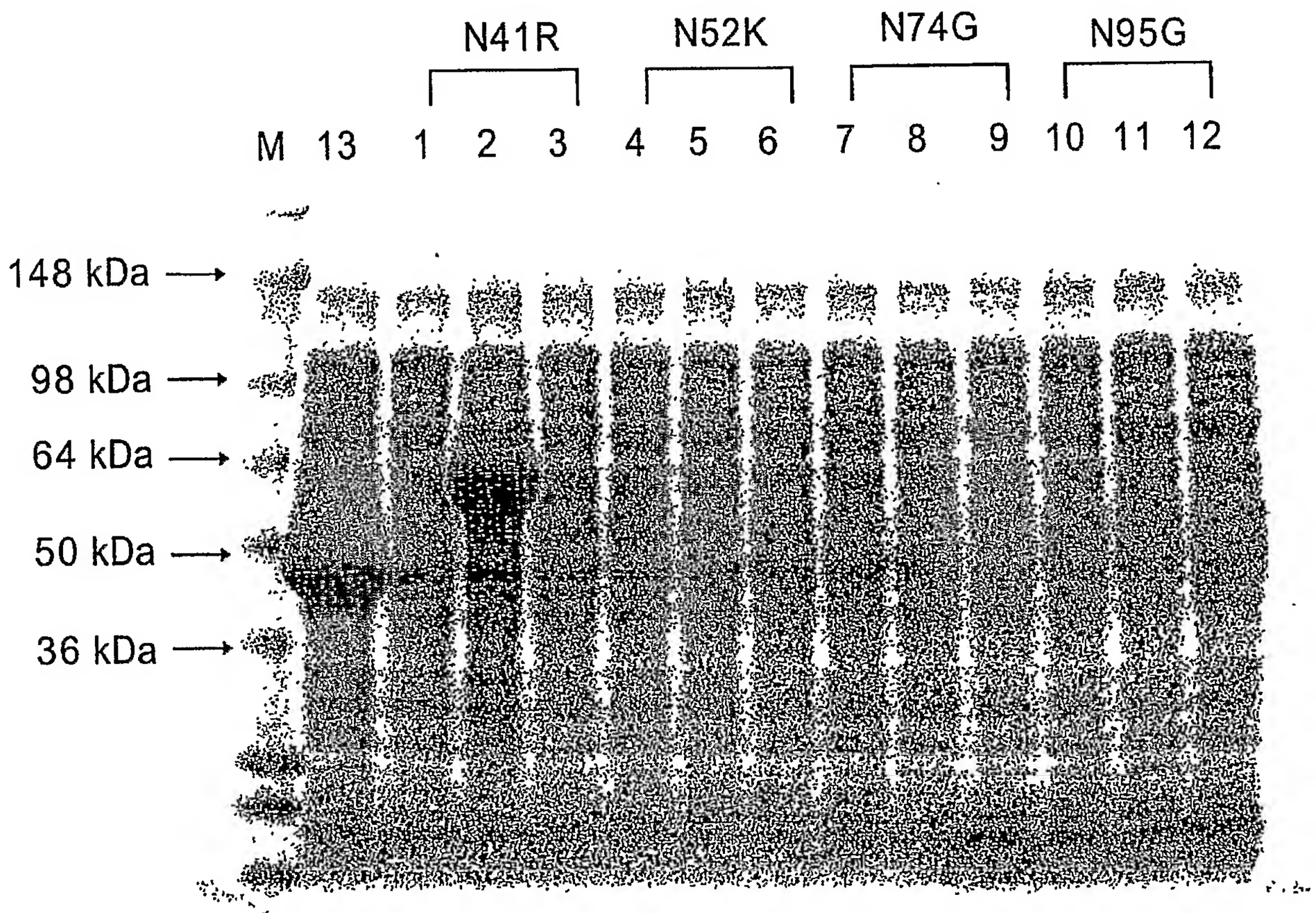

N41R
N52K
N74G
N95G
M 13 1 2 3 4 5 6 7 8 9 10 11 12
148 kDa
98 kDa
64 kDa
50 kDa
36 kDa

FIG. 5

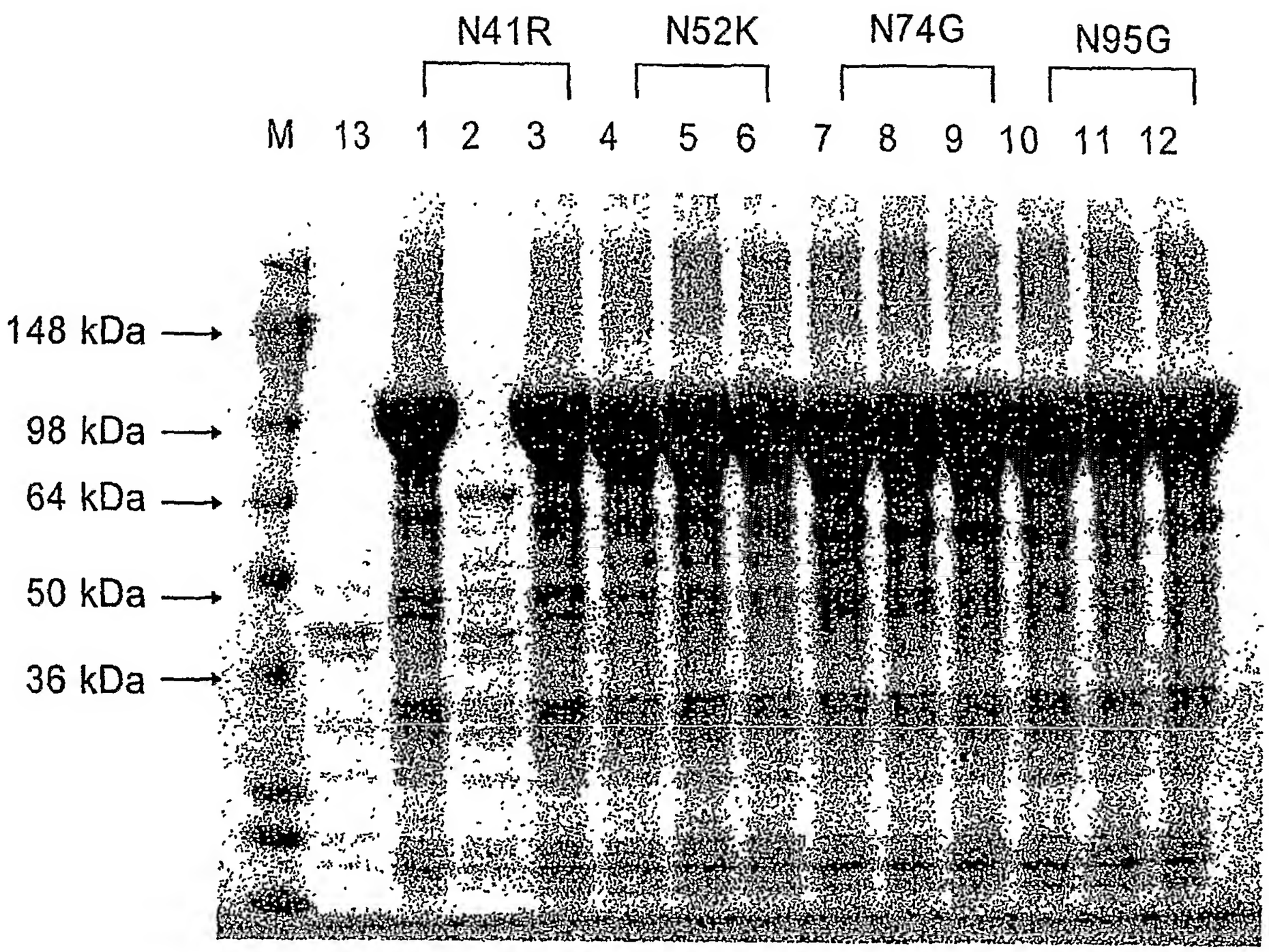

N41R
N52K
N74G
N95G
M 13 1 2 3 4 5 6 7 8 9 10 11 12
148 kDa
98 kDa
64 kDa
50 kDa
36 kDa

FIG. 6

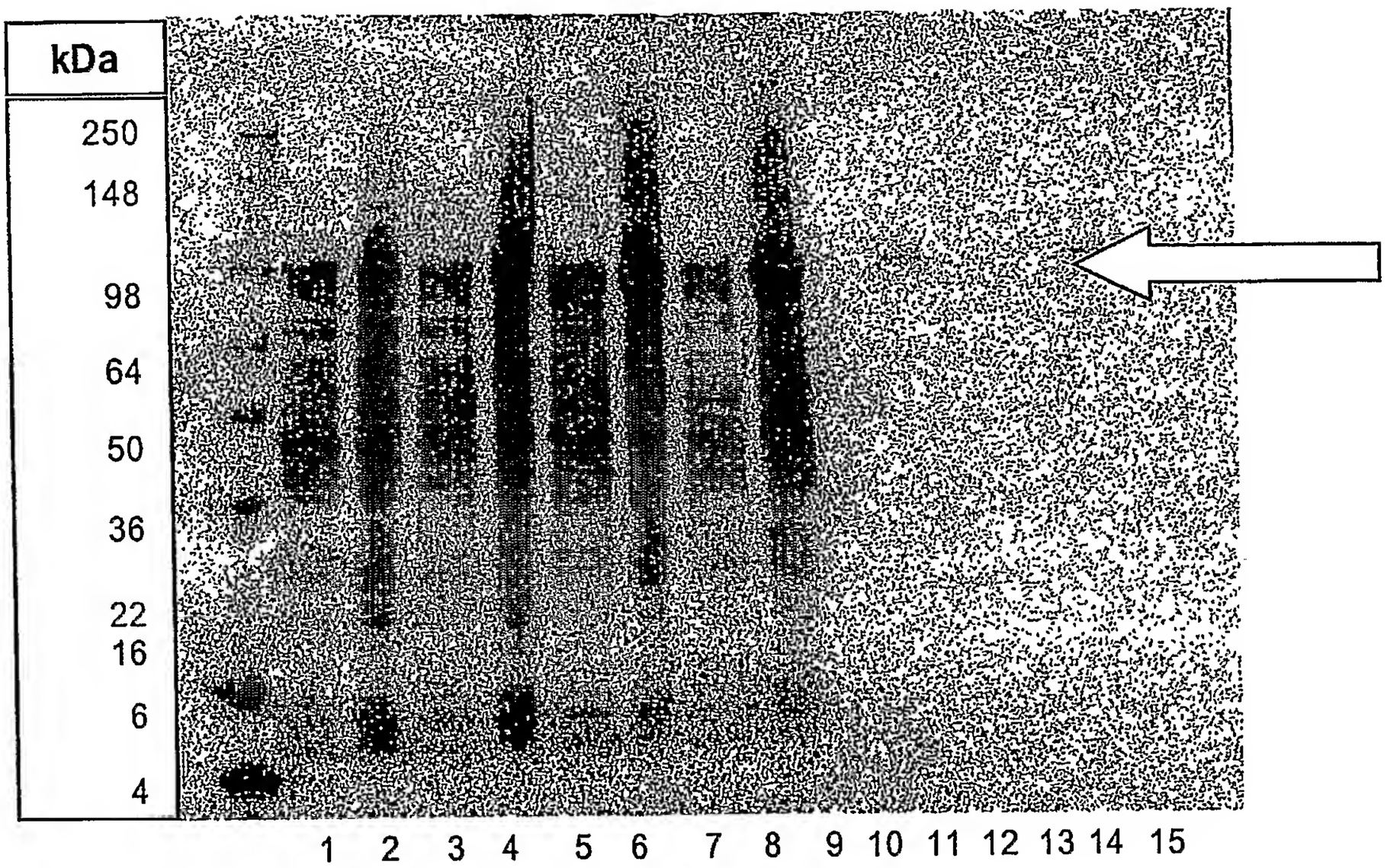


FIG. 7

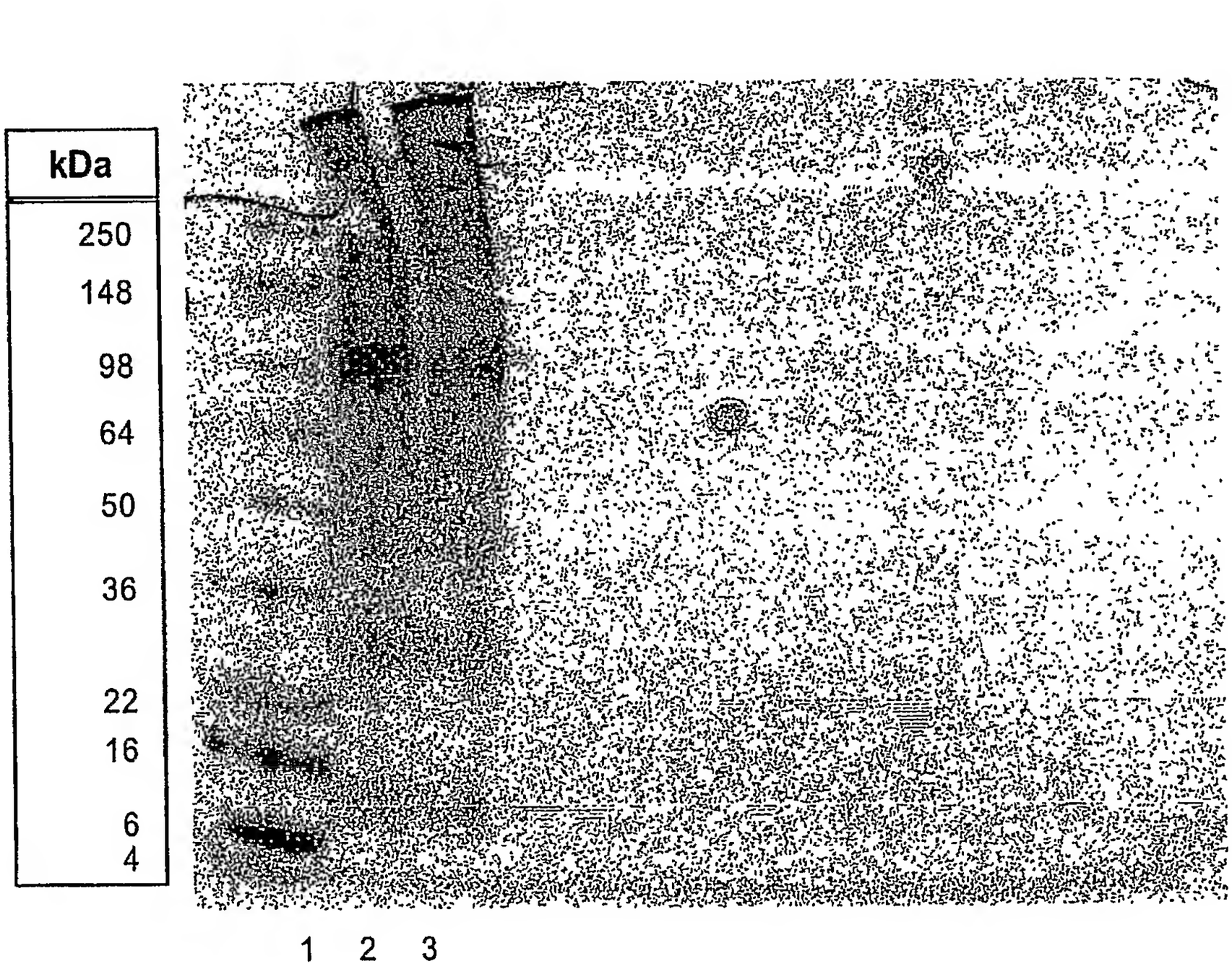


FIG. 8

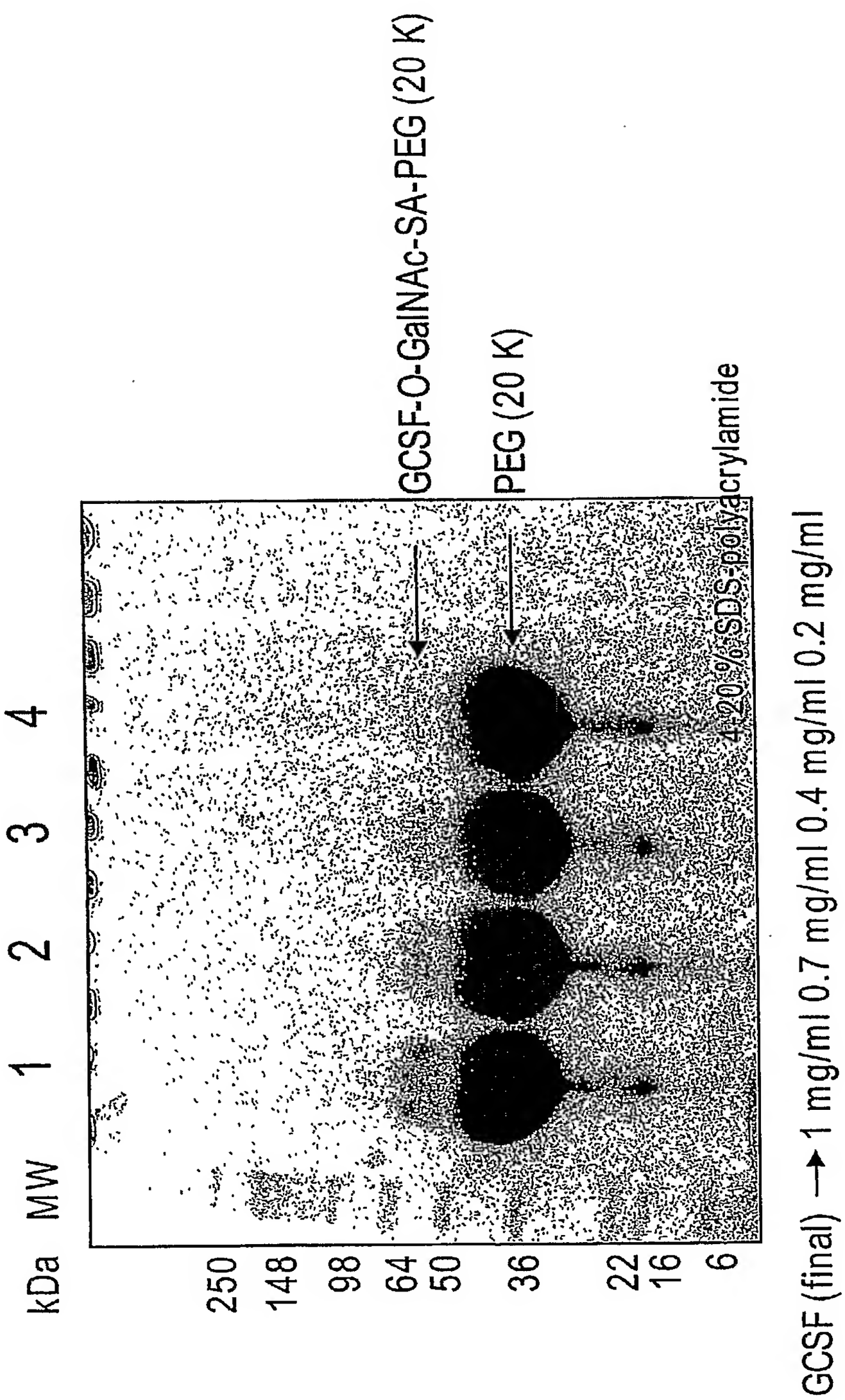

GCSF-O-GalNAc-SA-PEG (20 K)
PEG (20 K)
4-20% SDS-polyacrylamide
kDa MW 1 2 3 4
250
148
98
64
50
36
22
16
6
GCSF (final) → 1 mg/ml 0.7 mg/ml 0.4 mg/ml 0.2 mg/ml

FIG. 9

T2-41R

EooRi BamHi

```
   1  GAATTCGGAT CCAGGAAGGA GGACTGGAAT GAAATTGACC CCATTAAAAA GAAAGACCTT
      CTTAAGCCTA GGTCCTTCCT CCTGACCTTA CTTTAACTGG GGTAATTTTT CTTTCTGGAA

  61  CATCACAGCA ATGGAGAAGA GAAAGCACAA AGCATGGAGA CCCTCCCTCC AGGGAAAGTA
      GTAGTGTCGT TACCTCTTCT CTTTCGTGTT TCGTACCTCT GGGAGGGAGG TCCCTTTCAT
```

HindIII

```
 121  CGGTGGCCAG ACTTTAACCA GGAAGCTTAT GTTGGAGGGA CGATGGTCCG CTCCGGGCAG
      GCCACCGGTC TGAAATTGGT CCTTCGAATA CAACCTCCCT GCTACCAGGC GAGGCCCGTC
```

HindIII

```
 181  GACCGTTACG CCCGCAACAA GTTCAACCAG GTGGAGAGTG ATAAGCTTCG AATGGACAGA
      CTGGGAATGC GGGCGTTGTT CAAGTTGGTC CACCTCTCAC TATTCGAAGC TTACCTGTCT

 241  GCCATCCCTG ACACCCGGCA TGACCAGTGT CAGCGGAAGC AGTGGCGGGT GGATCTGCCG
      CGGTAGGGAC TGTGGGCCGT ACTGGTCACA GTCGCCTTCG TCACCGCCCA CCTAGACGGC

 301  GCCACCAGCG TGTTGATCAC GTTTCACAAT GAAGCCAGGT CGGCCCTACT CAGGACCGTG
      CGGTGGTCGC ACCACTAGTG CAAAGTGTTA CTTCGGTCCA GCCGGGATGA GTCCTGGCAC

 361  GTCAGGGTGC TTAAGAAAAG CCCGCCCCAT CTCATAAAAG AAATCATCTT GTTGGATGAC
      CAGTCGCACG AATTCTTTTC GGGCGGGGTA GAGTATTTTC TTTAGTAGAA CCACCTACTG

 421  TACAGCAATG ATCCTGAGGA CGGGGCTCTC TTGGGGAAAA TTGAGAAAGT GCGAGTTCTT
      ATGTCGTTAC TAGGACTCCT GCCCCGAGAG AACCCCTTTT AACTCTTTCA CGCTCAAGAA

 481  AGAAATGATC GACGAGAAGG CCTCATGCGC TCACGGGTTC GGGGGGCCGA TGCTGCCCAA
      TCTTTACTAG CTGCTCTTCC GGAGTACGCG AGTGCCCAAG CCCCCCGGCT ACGACGGGTT

 541  GCCAAGGTCC TGACCTTCCT GGACAGTCAC TGCGAGTGTA ATGAGCACTG GCTGGAGCCC
      CGGTTCCAGG ACTGGAAGGA CCTGTCAGTG ACGCTCACAT TACTCGTGAC CGACCTCGGG
```

ClaI

```
 601  CTCCTGGAAA GGGTGGCGGA GGACAGGACT CGGGTTGTGT CACCCATCAT CGATGTCATT
      GAGGACCTTT CCCACCGCCT CCTGTCCTGA GCCCAACACA GTGGGTAGTA GCTACAGTAA

 661  AATATGGACA ACTTTCAGTA TGTGGGGGCA TCTGCTGACT TGAAGGGCGG TTTTGATTGG
      TTATACCTGT TGAAAGTCAT ACACCCCCGT AGACGACTGA ACTTCCCGCC AAAACTAACC

 721  AACTTGGTAT TCAAGTGGGA TTACATGACG CCTGAGCAGA GAAGGTCCCG GCAGGGGAAC
      TTGAACCATA AGTTCACCCT AATGTACTGC GGACTCGTCT CTTCCAGGGC CGTCCCCTTG

 781  CCAGTCGCCC CTATAAAAAC CCCCATGATT GCTGGTGGGC TGTTTGTGAT GGATAAGTTC
      GGTCAGCGGG GATATTTTTG GGGGTACTAA CGACCACCCG ACAAACACTA CCTATTCAAG

 841  TATTTTGAAG AACTGGGGAA GTACGACATG ATGATGGATG TGTGGGGAGG AGAGAACCTA
      ATAAAACTTC TTGACCCCTT CATGCTGTAC TACTACCTAC ACACCCCTCC TCTCTTGGAT

 901  GAGATCTCGT TCCGCGTGTG GCAGTGTGGT GGCAGCCTGG AGATCATCCC GTGCAGCCGT
      CTCTAGAGCA AGGCGCACAC CGTCACACCA CCGTCGGACC TCTAGTAGGG CACGTCGGCA

 961  GTGGGACACG TGTTCCGGAA GCAGCACCCC TACACGTTCC CGGGTGGGAG TGGCACTGTC
      CACCCTGTGC ACAAGGCCTT CGTCGTGGGG ATGTGCAAGG GCCCACCGTC ACCGTGACAG

1021  TTTGCCCGAA ACACCCGCCG GGCAGCAGAG GTCTGGATGG ATGAATACAA AAATTTCTAT
      AAACGGGCTT TGTGGGCGGC CCGTCGTCTC CAGACCTACC TACTTATGTT TTTAAAGATA

1081  TATGCAGCAG TGCCTTCTGC TAGAAACGTT CCTTATGGAA ATATTCAGAG CAGATTGGAG
      ATACGTCGTC ACGGAAGACG ATCTTTGCAA GGAATACCTT TATAAGTCTC GTCTAACCTC
```

FIG. 10A

T2-41R

KpnI

1141	CTTAGGAAGA	AACTCAGCTG	CAAGCCTTTC	AAATGGTACC	TTGAAAATGT CTATCCAGAG
	GAATCCTTCT	TGAGTCGAC	GTTCGGAAAG	TTTACCATGG	AACTTTTACA GATAGGTCTC
1201	TTAAGGGTTC	CAGACCATCA	GGATATAGCT	TTTGGGGCCT	TGCAGCAGGG AACTAACTGC
	AATTCCCAAG	GTCTGGTAGT	CCTATATCGA	AAACCCCGGA	ACGTCGTCCC TTGATTGACG
1261	CTCGACACTT	TGGGACACTT	TGCTGATGGT	GTGGTTGGAG	TTTATGAATG TCACAATGCT
	GAGCTGTGAA	ACCCTGTGAA	ACGACTACCA	CACCAACCTC	AAATACTTAC AGTGTTACGA
1321	GGGGGAAACC	AGGAATGGGC	CTTGACGAAG	GAGAAGTCGG	TGAAGCACAT GGATTTGTGC
	CCCCCTTTGG	TCCTTACCCG	GAACTGCTTC	CTCTTCAGCC	ACTTCGTGTA CCTAAACACG
1381	CTTACTGTGG	TGGACCGGGC	ACCGGGCTCT	CTTATAAAGC	TGCAGGGCTG CCGAGAAAAT
	GAATGACACC	ACCTGGCCCG	TGGCCCGAGA	GAATATTTCG	ACGTCCCGAC GGCTCTTTTA
1441	GACAGCAGAC	AGAAATGGGA	ACAGATCGAG	GGCAACTCCA	AGCTGAGGCA CGTGGGCAGC
	CTGTCGTCTG	TCTTTACCCT	TGTCTAGCTC	CCGTTGAGGT	TCGACTCCGT GCACCCGTCG
1501	AACCTGTGCC	TGGACAGTCG	CACGGCCAAG	AGCGGGGGCC	TAAGCGTGGA GGTGTGTGGC
	TTGGACACGG	ACCTGTCAGC	GTGCCGGTTC	TCGCCCCCGG	ATTCGCACCT CCACACACCG

XhoI EooRI

1561	CCGGCCCTTT	CGCAGCAGTG	GAAGTTCACG	CTCAACCTGC	AGCAGTAGCT CGAGGAATTC
	GGCCGGGAAA	GCGTCGTCAC	CTTCAAGTGC	GAGTTGGACG	TCGTCATCGA GCTCCTTAAG

FIG. 10B

T2-52K

EooRi BamHi HindIII HindIII ClaI

```
   1  GAATTCGGAT CCAAAAAGAA AGACCTTCAT CACAGCAATG GAGAAGAGAA AGCACAAAGC
      CTTAAGCCTA GGTTTTTCTT TCTGGAAGTA GTGTCGTTAC CTCTTCTCTT TCGTGTTTCG

  61  ATGGAGACCC TCCCTCCAGG GAAAGTACGG TGGCCAGACT TTAACCAGGA AGCTTATGTT
      TACCTCTGGG AGGGAGGTCC CTTTCATGCC ACCGGTCTGA AATTGGTCCT TCGAATACAA

 121  GGAGGGACGA TGGTCCGCTC CGGGCAGGAC CCTTACGCCC GCAACAAGTT CAACCAGGTG
      CCTCCCTGCT ACCAGGCGAG GCCCGTCCTG GGAATGCGGG CGTTGTTCAA GTTGGTCCAC

 181  GAGAGTGATA AGCTTCGAAT GGACAGAGCC ATCCCTGACA CCCGGCATGA CCAGTGTCAG
      CTCTCACTAT TCGAAGCTTA CCTGTCTCGG TAGGGACTGT GGGCCGTACT GGTCACAGTC

 241  CGGAAGCAGT GGCGGGTGGA TCTGCCGGCC ACCAGCGTGG TGATCACGTT TCACAATGAA
      GCCTTCGTCA CCGCCCACCT CTGGCACCAG TGGTCGCACC ACTAGTGCAA AGTGTTACTT

 301  GCCAGGTCGG CCCTACTCAG GACCGTGGTC AGCGTGCTTA AGAAAAGCCC GCCCCATCTC
      CGGTCCAGCC GGGATGAGTC CTGGCACCAG TCGCACGAAT TCTTTTCGGG CGGGGTAGAG

 361  ATAAAAGAAA TCATCTTGGT GGATGACTAC AGCAATGATC CTGAGGACGG GGCTCTCTTG
      TATTTTCTTT AGTAGAACCA CCTACTGATG TCGTTACTAG GACTCCTGCC CCGAGAGAAC

 421  GGGAAAATTG ATCCTGAGGA AGTTCTTAGA AATGATCGAC GAGAAGGCCT CATGCGCTCA
      CCCTTTTAAC TCTTTCACGC TCAAGAATCT TTACTAGCTG CTCTTCCGGA GTACGCGAGT

 481  CGGGTTCGGG GGGCCGATGC TGCCCAAGCC AAGGTCCTGA CCTTCCTGGA CAGTCACTGC
      GCCCAAGCCC CCCGGCTACG CCTCGGGGAG TTCCAGGACT GGAAGGACCT GTCAGTGACG

 541  GAGTGTAATG AGCACTGGCT GGAGCCCCTC CTGGAAAGGG TGGCGGAGGA CAGGACTCGG
      CTCACATTAC TCGTGACCGA CCTCGGGGAG GACCTTTCCC ACCGCCTCCT GTCCTGAGCC

 601  GTTGTGTCAC CCATCATCGA TGTCATTAAT ATGGACAACT TTCAGTATGT GGGGGCATCT
      CAACACAGTG GGTAGTAGCT ACAGTAATTA TACCTGTTGA AAGTCATACA CCCCCGTAGA

 661  GCTGACTTGA AGGGCGGTTT TGATTGGAAC TTGGTATTCA AGTGGGATTA CATGACGCCT
      CGACTGAACT TCCCGCCAAA ACTAACCTTG AACCATAAGT TCACCCTAAT GTACTGCGGA

 721  GAGCAGAGAA GGTCCCGGCA GGGGAACCCA GTCGCCCCTA TAAAAACCCC CATGATTGCT
      CTCGTCTCTT CCAGGGCCGT CCCCTTGGGT CAGCGGGGAT ATTTTTGGGG GTACTAACGA

 781  GGTGGGCTGT TTGTGATGGA TAAGTTCTAT TTTGAAGAAC TGGGGAAGTA CGACATGATG
      CCACCCGACA AACACTACCT ATTCAAGATA AAACTTCTTG ACCCCTTCAT GCTGTACTAC

 841  ATGGATGTGT GGGGAGGAGA GAACCTAGAG ATCTCGTTCC GCGTGTGGCA GTGTGGTGGC
      TACCTACACA CCCCTCCTCT CTTGGATCTC TAGAGCAAGG CGCACACCGT CACACCACCG

 901  AGCCTGGAGA TCATCCCGTG CAGCCGTGTG GGACACGTGT TCCGGAAGCA GCACCCCTAC
      TGCAAGGGCC CACCGTCACC GTCGGCACAC CCTGTGCACA AGGCCTTCGT CGTGGGGAGT

 961  ACGTTCCCGG GTGGCAGTGG CACTGTCTTT GCCCGAAACA CCCGCCGGGC AGCAGAGGTC
      TGCAAGGGCC CACCGTCACC GTCGGCACAC CGGGCTTTGT GGGCGGCCCG TCGTCTCCAG

1021  TGGATGGATG AATACAAAAA TTTCTATTAT GCAGCAGTGC CTTCTGCTAG AAACGTTCCT
      ACCTACCTAC TTATGTTTTT AAAGATAATA CGTCGTCACG GAAGACGATC TTTGCAAGGA

1081  TATGGAAATA TTCAGAGCAG ATTGGAGCTT AGGAAGAAAC TCAGCTGCAA GCCTTTCAAA
      ATACCTTTAT AAGTCTCGTC TAACCTCGAA TCCTTCTTTG AGTCGACGTT CGGAAAGTTT
```

FIG. 11A

T2-52K

KpnI

```
1141  TGGTACCTTG AAAATGTCTA TCCAGAGTTA AGGGTTCCAG ACCATCAGGA TATAGCTTTT
      ACCATGGAAC TTTTACAGAT AGGTCTCAAT TCCCAAGGTC TGGTAGTCCT ATATCGAAAA

1201  GGGGCCTTGC AGCAGGGAAC TAACTGCCTC GACACTTTGG GACACTTTGC TGATGGTGTG
      CCCCGGAACG TCGTCCCTTG ATTGACGGAG CTGTGAAACC CTGTGAAACG ACTACCACAC

1261  GTTGGAGTTT ATGAATGTCA CAATGCTGGG GCAAACCAGG AATGGGCCTT GACGAAGGAG
      CAACCTCAAA TACTTACAGT GTTACGACCC CCTTTGGTCC TTACCCGGAA CTGCTTCCTC

1321  AAGTCGGTGA AGCACATGGA TTTGTGCCTT ACTGTGGTGG ACCGGGCACC GGGCTCTCTT
      TTCAGCCACT TCGTGTACCT AAACACGGAA TGACACCACC TGGCCCGTGG CCCGAGAGAA

1381  ATAAAGCTGC AGGGCTGCCG AGAAAATGAC AGCAGACAGA AATGGGAACA GATCGAGGGC
      TATTTCGACG TCCCGACGGC TCTTTTACTG TCGTCTGTCT TTACCCTTGT CTAGCTCCCG

1441  AACTCCAAGC TGAGGCACGT GGGCAGCAAC CTGTGCCTGG ACAGTCGCAC GGCCAAGAGC
      TTGAGGTTCG ACTCCGTGCA CCCGTCGTTG GACACGGACC TGTCAGCGTG CCGGTTCTCG

1501  GGGGGCCTAA GCGTGGAGGT GTGTGGCCCG GCCCTTTCGC AGCAGTGGAA GTTCACGCTC
      CCCCCGGATT CGCACCTCCA CACACCGGGC CGGGAAAGCG TCGTCACCTT CAAGTGCGAG
```

XhoI

KpnI

```
1561  AACCTGCAGC AGTAGCTCGA GGAATTC
      TTGGACGTCG TCATCGAGCT CCTTAAG
```

FIG. 11B

T2-74G

EooRi BamHi HindIII

```
   1  GAATTCGGAT CCGGGAAAGT ACGGTGGCCA GACTTTAACC AGGAAGCTTA TGTTGGAGGG
      TGCTACCAGG GGCCCTTTCA TGCCACCGGT CTGAAATTGG TCCTTCGAAT ACCACCTCCC

  61  ACGATGGTCC GCTCCGGGCA GGACCCTTAC GCCCGCAACA AGTTCAACCA GGTGGAGAGT
      TGCTACCAGG CGAGGCCCGT CCTGGGAATG CGGGCGTTGT TCAAGTTGGT CCACCTCTCA
```

HindIII

```
 121  GATAAGCTTC GAATGGACAG AGCCATCCCT GACACCCGGC ATGACCAGTG TCAGCGGAAG
      CTATTCGAAG CTTACCTGTC TCGGTAGGGA CTGTGGGCCG TACTGGTCAC AGTCGCCTTC

 181  CAGTGGCGGG TGGATCTGCC GGCCACCAGC GTGGTGATCA CGTTTCACAA TGAAGCCAGG
      GTCACCGCCC ACCTAGACGG CCGGTGGTCG CACCACTAGT GCAAAGTGTT ACTTCGGTCC

 241  TCGGCCCTAC TCAGGACCGT TCTGCCGGCC CTTAAGAAAA GCCCGCCCCA TCTCATAAAA
      AGCCGGGATG AGTCCTGGCA CTGGCACCAG GAATTCTTTT CGGGCGGGGT AGAGTATTTT

 301  GAAATCATCT TGGTGGATGA CTACAGCAAT GATCCTGAGG ACGGGGCTCT CTTGGGGAAA
      CTTTAGTAGA ACCACCTACT GATGTCGTTA CTAGGACTCC TGGCCCGAGA GAACCCCTTT

 361  ATTGAGAAAG TGCGAGTTCT AGCCAAGGTC CGACGACAAG GCCTCATGCG CTCACGGGTT
      TAACTCTTTC ACGCTCAAGA ATCTTTACTA GCTGCTCTTC CGGAGTACGC GAGTGCCCAA

 421  CGGGGGGCCG ATGCTGCCCA AGCCAAGGTC CTGACCTTCC TGGACAGTCA CTGCGAGTGT
      GCCCCCCGGC TACGACGGGT TCGGTTCCAG GACTGGAAGG ACCTGTCAGT GACGCTCACA

 481  AATGAGCACT GGCTGGAGCC CCTCCTGGAA AGGGTGGCGG AGGACAGGAC TCGGGTTGTG
      TTACTCGTGA CCGACCTCGG GGAGGACCTT TCCCACCGCC TCCTGTCCTG AGCCCAACAC
```

ClaI

```
 541  TCACCCATCA TCGATGTCAT TAATATGGAC AACTTTCAGT ATGTGGGGGC ATCTGCTGAC
      AGTGGGTAGT AGCTACAGTA ATTATACCTG TTGAAAGTCA TACACCCCCG TAGACGACTG

 601  TTGAAGGGCG GTTTTGATTG GAACTTGGTA TTCAAGTGGG ATTACATGAC GCCTGAGCAG
      AACTTCCCGC CAAAACTAAC CTTGAACCAT AAGTTCACCC TAATGTACTG CGGACTCGTC

 661  AGAAGGTCCC GGCAGGGGAA CCCAGTCGCC CCTATAAAAA CCCCCATGAT TGCTGGTGGG
      TCTTCCAGGG CCGTCCCCTT GGGTCAGCGG GGATATTTTT GGGGGTACTA ACGACCACCC

 721  CTGTTTGTGA TGGATAAGTT CTATTTTGAA GAACTGGGGA AGTACGACAT GATGATGGAT
      GACAAACACT ACCTATTCAA GATAAAACTT CTTGACCCCT TCATGCTGTA CTACTACCTA

 781  GTGTGGGGAG GAGAGAACCT AGAGATCTCG TTCCGCGTGT GGCAGTGTGG TGGCAGCCTG
      CACACCCCTC CTCTCTTGGA TCTCTAGAGC AAGGCGCACA CCGTCACACC ACCGTCGGAC

 841  GAGATCATCC CGTGCAGCCG TGTGGGACAC GTGTTCCGGA AGCAGCACCC CTACACGTTC
      CTCTAGTAGG GCACGTCGGC ACACCCTGTG CACAAGGCCT TCGTCGTGGG GATGTGCAAG

 901  CCGGGTGGCA GTGGCACTGT CTTTGCCCGA AACACCCGCC GGGCAGCAGA GGTCTGGATG
      GGCCCACCGT CACCGTGACA GAAACGGGCT TTGTGGGCGG CCCGTCGTCT CCAGACCTAC

 961  GATGAATACA AAAATTTCTA TTATGCAGCA GTGCCTTCTG CTAGAAAGCT AGCAGAGGTC
      CTACTTATGT TTTTAAAGAT AATACGTCGT CACGGAAGAC GGGCGGCCCG AGGAATACCT
```

KpnI

```
1021  AATATTCAGA GCAGATTGGA GCTTAGGAAG AAACTCAGCT GCAAGCCCTT CAAATGGTAC
      TTATAAGTCT CGTGTAACCT CGAATCCTTC TTTGAGTCGA CGTTCGGAAA GTTTACCATG
```

Kpn

```
1081  CTTGAAAATG TCTATCCAGA GTTAAGGGTT CCAGACCATC AGGATATAGC TTTTGGGGCC
      GAACTTTTAC AGATAGGTCT CAATTCCCAA GGTCTGGTAG TCCTATATCG AAAACCCCGG
```

FIG. 12A

T2-74G

1141 TTGCAGGAGG GAACTAACTG CCTCGACACT TTGGGACACT TTGCTGATGG TGTGGTTGGA
AACGTCGTCC CTTGATTGAC GGAGCTGTGA AACCCTGTGA AACGACTACC ACACCAACCT

1201 GTTTATGAAT GTCACAATGC TGGGGGAAAC CAGGAATGGG CCTTGACGAA GGAGAAGTCG
CAAATACTTA CAGTGTTACG ACCCCCTTTG GTCCTTACCC GGAACTGCTT CCTCTTCAGC

1261 GTGAAGCACA TGGATTTGTG CCTTACTGTG GTGGACCGGG CACCGGGCTC TCTTATAAAG
CACTTCGTGT ACCTAAACAC GGAATGACAC CACCTGGCCC GTGGCCCGAG CCCGTTGAGG

1321 CTGCAGGGCT TGGATTTGTG TGACAGCAGA CAGAAATGGG AACAGATCGA GGGCAACTCC
GACGTCCCGA ACCTAAACAC ACTGTCGTCT GTCTTTACCC TTGTCTAGCT CCCGTTGAGG

1381 AAGCTGAGGC ACGTGGGCAG CAACCTGTGC CTGGACAGTC GCACGGCCAA GAGCGGGGGC
TTCGACTCCG TGCACCCGTC GTTGGACACG GACCTGTCAG CGTGCCGGTT CTCGCCCCCG

1441 CTAAGCGTGG AGGTGTGTGG CCCGGCCCTT TCGCAGCAGT GGAAGTTCAC GCTCAACCTG
GATTCGCACC TCCACACACC GGGCCGGGAA ACGCTCGTCA CCTTCAAGTG CGAGTTGGAC

XhoI
EccRI

1501 CAGCAGTAGC TCGAGGAATT C
GTCGTCATCG AGCTCCTTAA G

FIG. 12B

T2-94

BamHi
EooRi

```
1    GAATTCGGAT CCGGGCAGGA CCCTTACGCC CGCAACAAGT TCAACCAGGT GGAGAGTGAT
     CTTAAGCCTA GGCCCGTCCT GGGAATGCGG GCGTTGTTCA AGTTGGTCCA CCTCTCACTA

          HindIII
61   AAGCTTCGAA TGGACAGAGC CATCCCTGAC ACCCGGCATG ACCAGTGTCA GCGGAAGCAG
     TTCGAAGCTT ACCTGTCTCG GTAGGGACTG TGGGCCGTAC TGGTCACAGT CGCCTTCGTC

121  TGGCGGGTGG ATCTGCCGGA CACCAGCGTG GTGATCACGT TTCACAATGA AGCCAGGTCG
     ACCGCCCACC TAGACGGCCG GTGGTCGCAC CACTAGTGCA AAGTGTTACT TCGGTCCAGC

181  GCCCTACTCA GGACCGTGGT CAGCGTGCTT AAGAAAAGCC CGCCCCATCT CATAAAAGAA
     CGGGATGAGT CCTGGCACCA GTCGCACGAA TTCTTTTCGG GCGGGGTAGA GTATTTTCTT

241  ATCATCTTGG TGGATGACTA CAGCAATGAT CCTGAGGACG GGGCTCTCTT GGGGAAAATT
     TAGTAGAACC ACCTACTGAT GTCGTTACTA GGACTCCTGC CCCGAGAGAA CCCCTTTTAA

301  GAGAAAGTGC GAGTTCTTAG AAATGATCGA CGAGAAAGGC TCATGCGCTC ACGGGTTCGG
     CTCTTTCACG CTCAAGAATC TTTACTAGCT GCTCTTCCGG AGTACGCGAG TGCCCAAGCC

361  GGGGCCGATG TGGAGCCCCT CAAGGTCCTG ACCTTCCTGG ACAGTCACTG CGAGTGTAAT
     CCCCGGCTAC GACGGGTTCG GACGGGTTCG TGGAAGGACC TGTCAGTGAC GCTCACATTA

421  GAGCACTGGC TGGAGCCCCT CCTGGAAAGG GTGGCGGAGG ACAGGACTCG GGTTGTGTCA
     CTCGTGACCG ACCTCGGGGA GGACCTTTCC CACCGCCTCC TGTCCTGAGC CCAACACAGT

          ClaI
481  CCCATCATCG ATGTCATTAA TATGGACAAC TTTCAGTATG TGGGGGCATC TGCTGACTTG
     GGGTAGTAGC TACAGTAATT ATACCTGTTG AAAGTCATAC ACCCCCGTAG ACGACTGAAC

541  AAGGGCGGTT TTGATTGGAA CTTGGTATTC AAGTGGGATT ACATGACGCC TGAGCAGAGA
     TTCCCGCCAA AACTAACCTT GAACCATAAG TTCACCCTAA TGTACTGCGG ACTCGTCTCT

601  AGGTCCCGGC AGGGGAACCC AGTCGCCCCT ATAAAAACCC CCATGATTGC TGGTGGGCTG
     TCCAGGGCCG TCCCCTTGGG TCAGCGGGGA TATTTTTGGG GGTACTAACG ACCACCCGAC

661  TTTGTGATGG ATAAGTTCTA TTTTGAAGAA CTGGGGAAGT ACGACATGAT GATGGATGTG
     AAACACTACC TATTCAAGAT AAAACTTCTT GACCCCTTCA TGCTGTACTA CTACCTACAC

721  TGGGGAGGAG AGAACCTAGA GATCTCGTTC CGCGTGTGGC AGTGTGGTGG CAGCCTGGAG
     ACCCCTCCTC TCTTGGATCT CTAGAGCAAG GCGCACACGC TCACACCACC GTCGGACCTC

781  ATCATCCCGT GCAGCCGTGT GGGACACGTG TTCCGGAAGC AGCACCCCTA CACGTTCCCG
     TAGTAGGGCA CGTCGGCACA CCCTGTGCAC AAGGCCTTCG TCGTGGGGAT GTGCAAGGGC

841  GGTGGCAGTG GCACTGTCTT TGCCCGAAAC ACCCGCCGGG CAGCAGAGGT CTGGATGGAT
     CCACCGTCAC CGTGACAGAA CTTGGATCTC TGGGCGGCCC GTCGTCTCCA GACCTACCTA

901  CAATACAAAA ATTTCTATTA TGCAGCAGTG CCTTCTGCTA GAAACGTTCC TTATGGAAAT
     CTTATGTTTT TAAAGATAAT ACGTCGTCAC GGAAGACGAT TCGGAAAGTT AATACCTTTA

                                                          KpnI
961  ATTCAGAGCA GATTGGAGCT TAGGAAGAAA CTCAGCTGCA AGCCTTTCAA ATGGTACCTT
     TAAGTCTCGT CTAACCTCGA ATCCTTCTTT GAGTCGACGT TCGGAAAGTT TACCATGGAA

1021 GAAAATGTCT ATCCAGAGTT AAGGGTTCCA GACCATCAGG ATATAGCTTT TGGGGCCTTG
     CTTTTACAGA TAGGTCTCAA TTCCCAAGGT CTGGTAGTCC TATATCGAAA ACCCCGGAAC

1081 CAGCAGGGAA CTAACTGCCT CGACACTTTG GGACACTTTG CTGTAGGTGT GGTTGGAGTT
     GTCGTCCCTT GATTGACGGA GCTGTGAAAC CCTGTGAAAC GACTACCACA CCAACCTCAA
```

FIG. 13A

T2-94

```
1141  TAGAATGTC  ACAATGCTGG GGGAAACCAG GAATGGGCCT TGACGAAGGA GAAGTCGGTG
      ATACTTACAG TGTTACGACC CCCTTTGGTC CTTAACCGGA ACTGCTTCCT CTTCAGCCAC

1201  AAGCACATGG ATTTGTGCCT TACTGTGGTG GACCGGGCAC CGGGCTCTCT TATAAAGCTG
      TTCGTGTACC TAAACACGGA ATGACACCAC CTGGCCCGTG GCCCGAGAGA ATATTTCGAC

1261  CAGGGCTGCC GAGAAAATGA CAGCAGACAG AAATGGGAAC AGATCGAGGG CAAGTCCAG
      GTCCCGACGG CTCTTTTACT TTTACCCTTG TTTACCCTTG TCTAGCTCCC GTTGAGTTC

1321  CTGAGGCACG TGGGCAGCAA CCTGTGCCTG GACAGTCGCA CGGCCAAGAG CGGGGGCCTA
      GACTCCGTGC ACCCGTCGTT GGACACGGAC CTGTCAGGGT GCCGGTTCTC GCCCCCGGAT

1381  AGCGTGGAGG TGTGTGGCCC GGCCCTTTCG CAGCAGTGGA AGTTCACGCT CAACCTGCAG
      TCGCACCTCC ACACACCGGG CCGGGAAAGC GTCGTCACCT TCAAGTGCGA GTTGGACGTC

                XhoI
                    EccRI
1441  CAGTAGCTCG AGGAATTC
      GTCATCGAGC TCCTTAAG
```

FIG. 13B

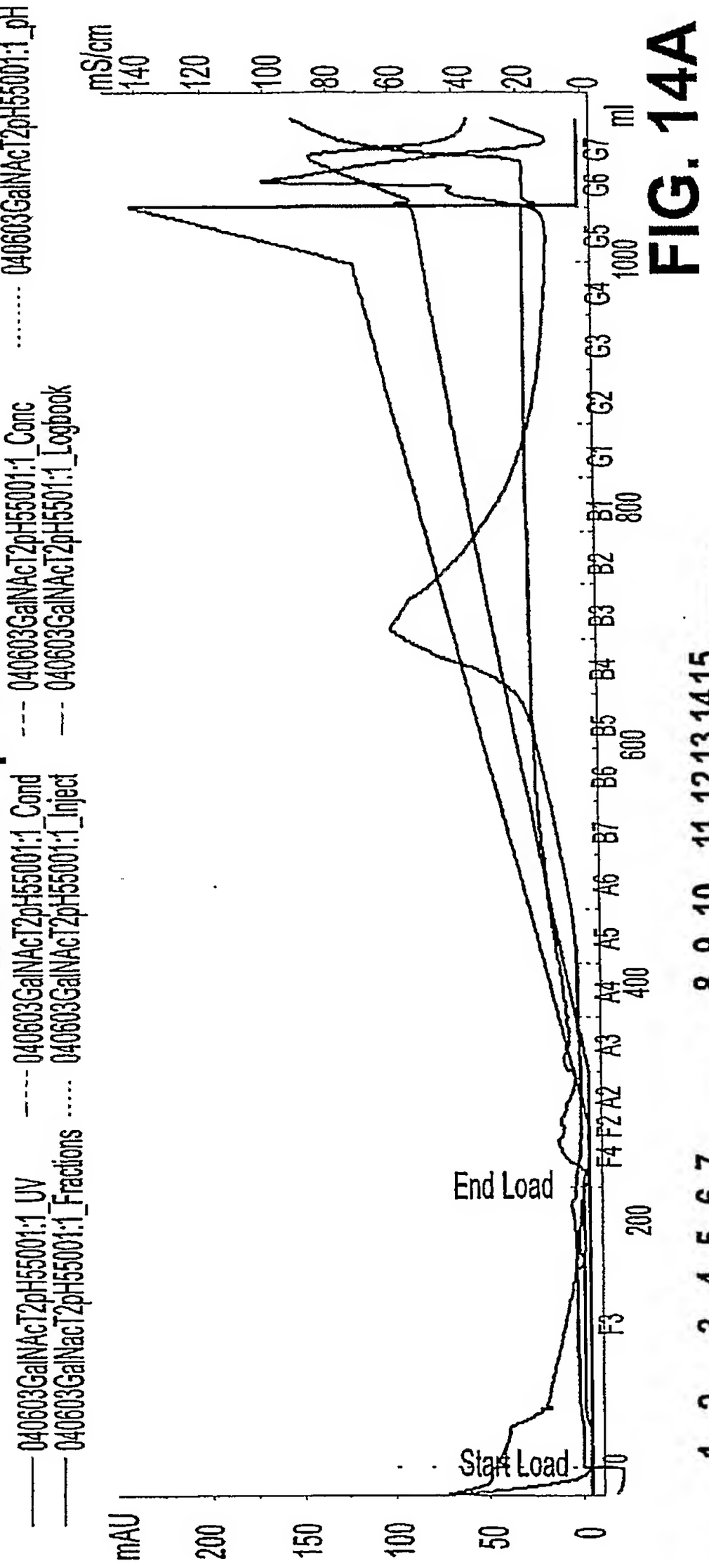


FIG. 14A

FIG. 14B

AC#	Sample Description	Activity (U/L)	%RSD	Volume (ml)	Activity (U)
AC04-08477	040603 1b A6	0.10	22.3	45	0.00
AC04-08478	040603 1b B7	0.20	42.4	45	0.01
AC04-08479	040603 1b B6	0.15	10.7	45	0.01
AC04-08480	040603 1b B5	0.19	78.4	45	0.01
AC04-08481	040603 1b B4	0.06	0.0	45	0.00
AC04-08482	040603 1b B3	0.10	4.9	45	0.00
AC04-08483	040603 1b B2	0.39	76.4	45	0.02
AC04-08484	040603 1b C6	0.02	35.1		0.00
AC04-08485	040603 1b FR	0.04	2.2	1000	0.04
AC04-08486	040603 1b cFR	0.09	2.3	250	0.02
AC04-08487	040603 1b CP	-0.01	8.4	750	0.00
AC04-08488	040603 1b DP	-0.02	9.0	1000	-0.02
AC04-08489	040603 1b QL	0.03	0.6	245	0.01
AC04-08490	040603 1b QFT	-0.04	8.3	245	-0.01
AC04-08491	040603 1b QW	-0.02	19.8	48	0.00

FIG. 14C

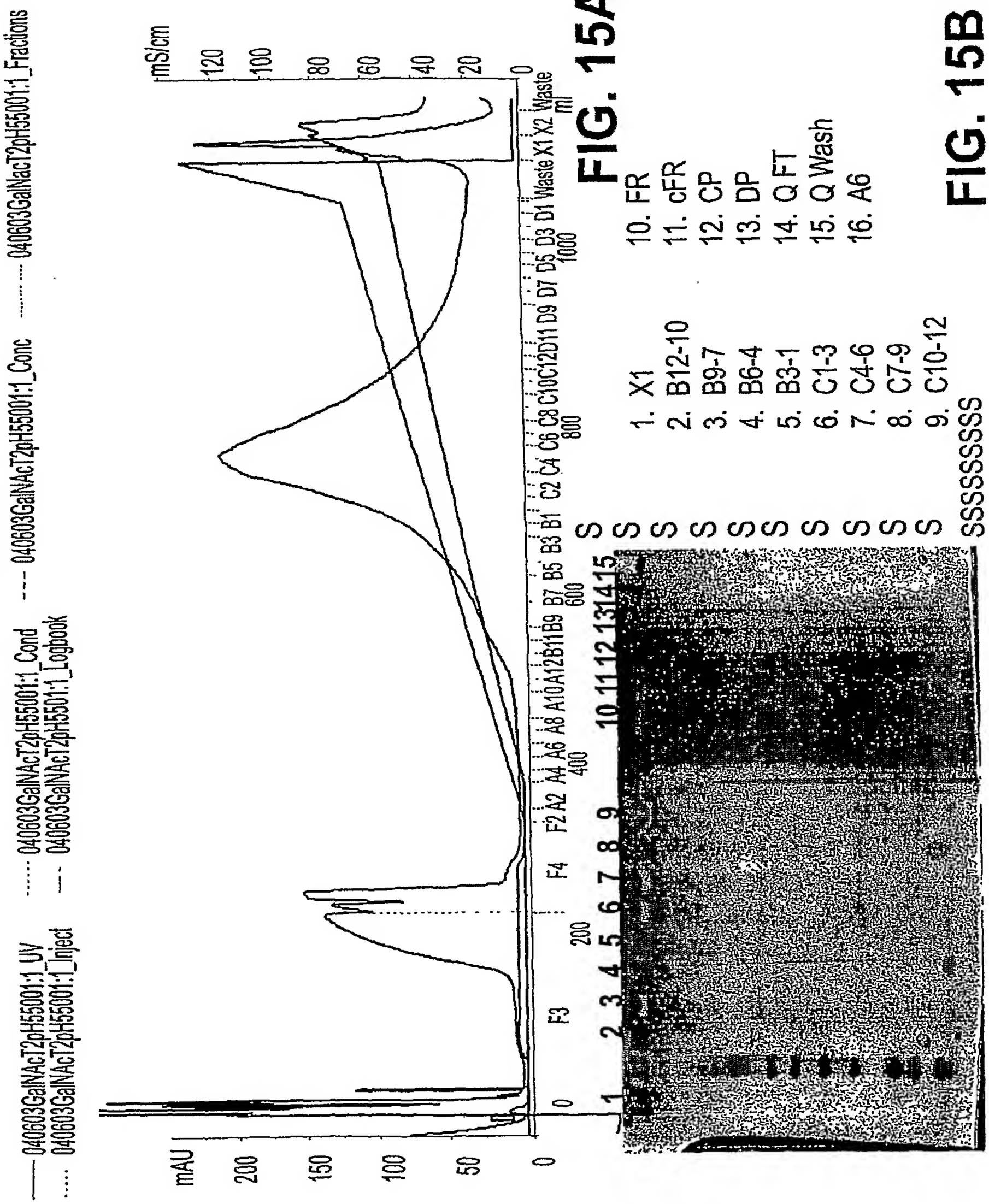


FIG. 15A

FIG. 15B

AC#	Sample Description	Activity (U/L)	%RSD	Volume (ml)	Activity (U)
Negative Control (No Enzyme)		0.00	14.6		
AC04-08462	040603 1a FR	1.08	24.0	1000	1.08
AC04-08463	040603 1a cFR	1.03	4.4	250	0.26
AC04-08464	040603 1a CP	0.13	17.8	750	0.09
AC04-08465	040603 1a DP	0.08	5.7	1000	0.08
AC04-08466	040603 1a QFT	0.07	34.0	270	0.02
AC04-08467	040603 1a QW	0.05	24.4	48	0.00
AC04-08468	040603 1a B12-10	1.60	9.0	45	0.07
AC04-08469	040603 1a B9-7	2.21	11.7	45	0.10
AC04-08470	040603 1a B6-4	0.88	17.0	45	0.04
AC04-08471	040603 1a B3-1	0.43	4.4	45	0.02
AC04-08472	040603 1a C1-3	0.36	22.2	45	0.02
AC04-08473	040603 1a C4-6	0.32	4.9	45	0.01
AC04-08474	040603 1a C7-9	0.22	1.3	45	0.01
AC04-08475	040603 1a C10-12	0.27	71.8	45	0.01
AC04-08476	040603 1a X1	0.04	6.3	29	0.00

FIG. 15C

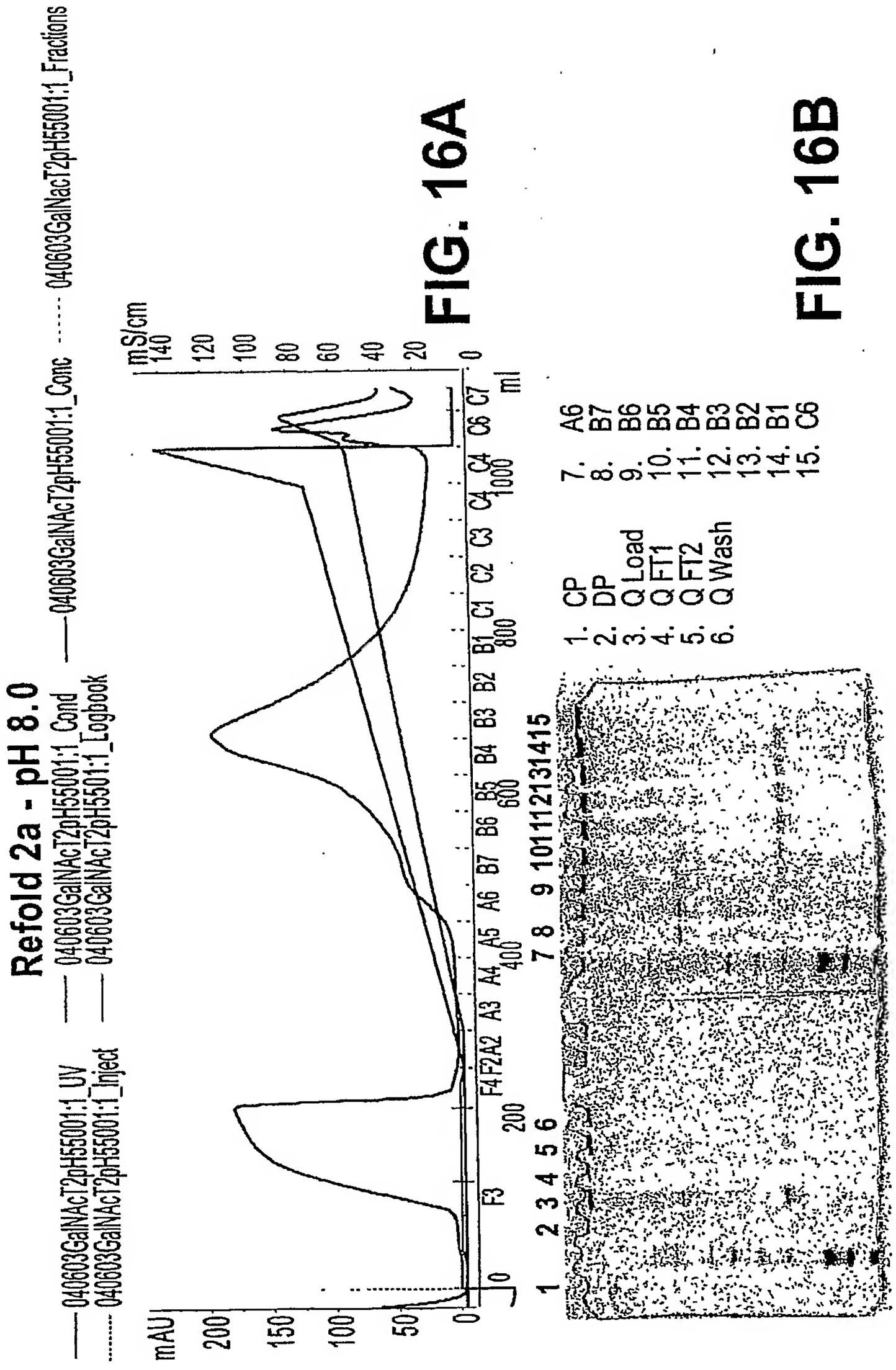


FIG. 16A

FIG. 16B

AC#	Sample Description	Activity (U/L)	%RSD	Volume (ml)	Activity (U)	A280	Conc (mg/mL)	Mass (mg)	Specif. Act. (U/mg)
AC04-08492	040603 2a A6	3.23	9.6	45	**0.15**	0.027	0.0179	0.806	0.186
AC04-08493	040603 2a B7	4.14	0.2	45	**0.19**	0.064	0.0424	1.908	0.099
AC04-08494	040603 2a B6	1.79	3.7	45	**0.08**	0.096	0.0636	2.862	0.028
AC04-08495	040603 2a B5	0.52	15.0	45	0.02	0.223	0.148	33.3	
AC04-08496	040603 2a B4	0.12	8.1	45	0.01				
AC04-08497	040603 2a B3	0.16	18.6	45	0.01				
AC04-08498	040603 2a B2	0.03	18.4	45	0.00				
AC04-08499	040603 2a B1	0.00	1.4	45	0.00				
AC04-08500	040603 2a C6	0.07	24.8	45	0.00				
AC04-08501	040603 2a CP	0.30	73.3	750	0.22				
AC04-08502	040603 2a DP	-0.05	1.26	1000	-0.05				
AC04-08503	040603 2a QL	1.24	5.0	265	0.33	0.651	0.431	114.2	
AC04-08504	040603 2a QFT1	-0.24	4.3	130	-0.03	0.000	-	-	
AC04-08505	040603 2a QFT2	-0.22	8.5	90	-0.02	0.284	0.188	16.92	
AC04-08506	040603 2a QW	-0.21	2.3	48	-0.01	0.068	0.045	2.16	

FIG. 16C

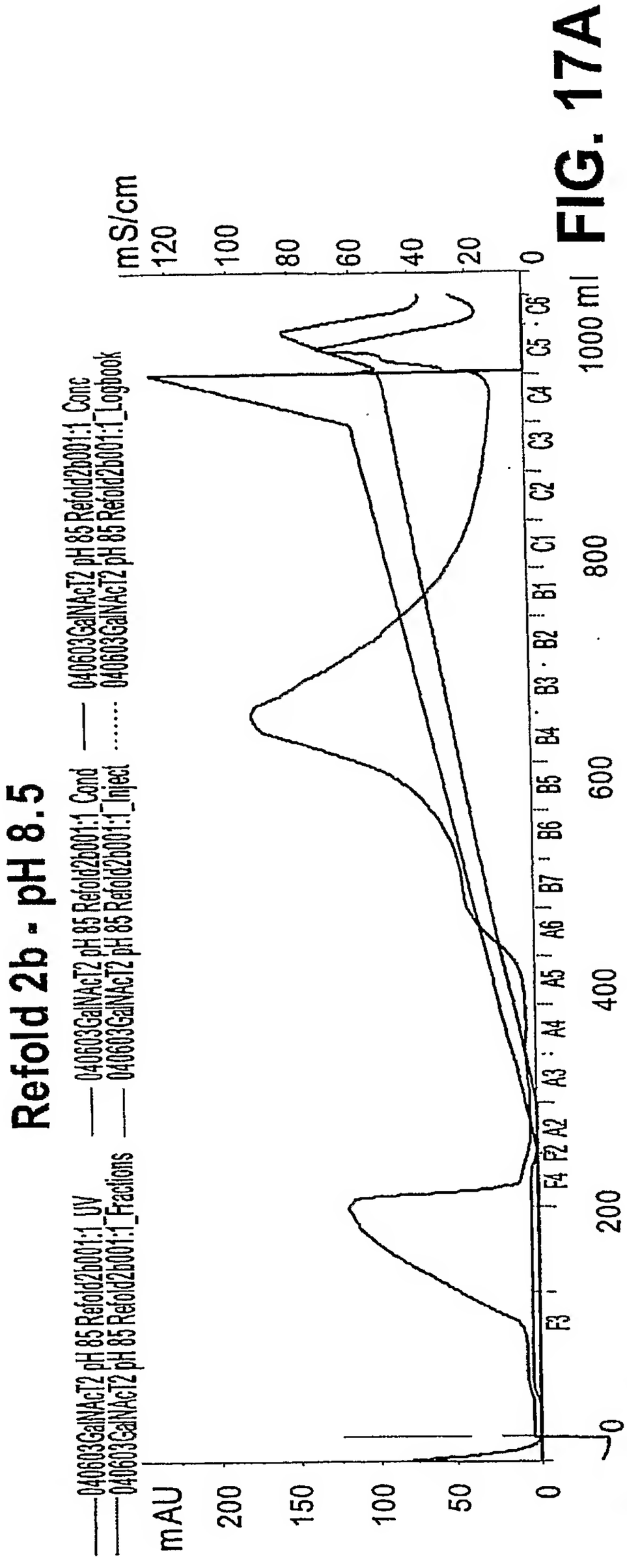


FIG. 17A

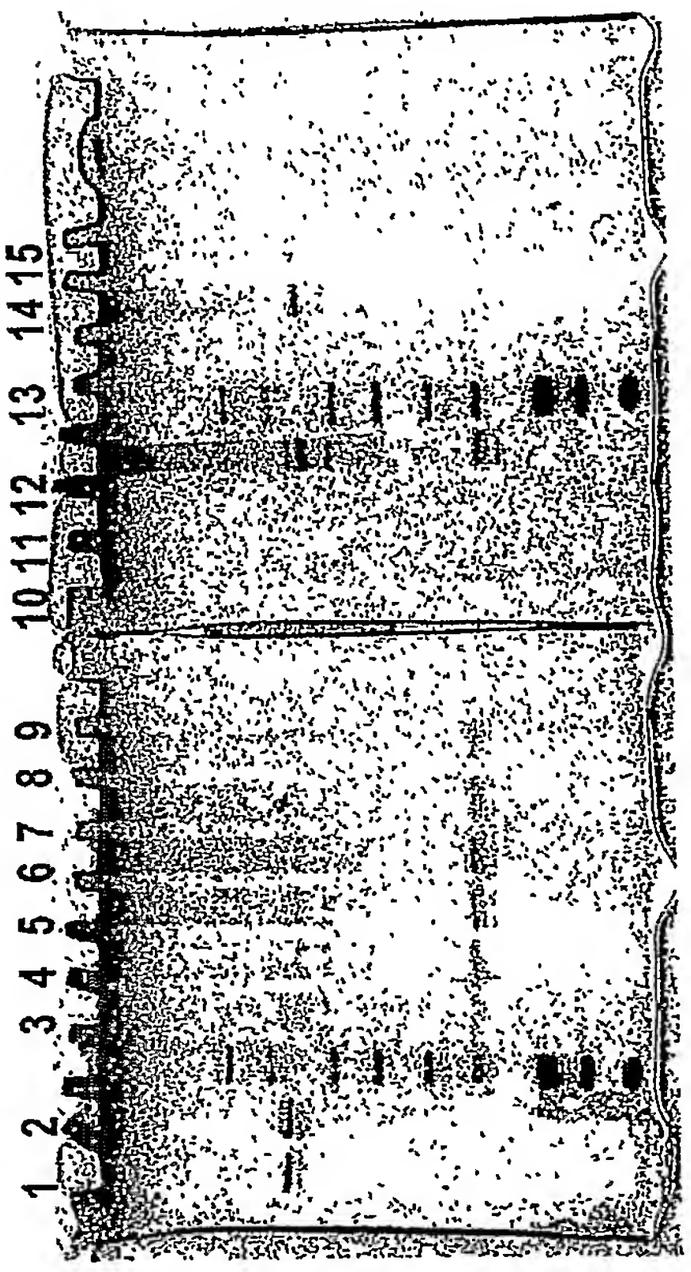


FIG. 17B

AC#	Sample Description	Activity (U/L)	%RSD	Volume (ml)	Activity (U)	A280	Conc (mg/mL)	Mass (mg)	Specif. Act. (U/mg)
AC04-08507	040603 2b A6	2.40	14.76	45	**0.11**	0.029	0.0192	0.864	0.127
AC04-08508	040603 2b B7	3.77	11.14	45	**0.17**	0.058	0.0384	1.728	0.098
AC04-08509	040603 2b B6	1.17	3.84	45	**0.05**				
AC04-08510	040603 2b B5	0.10	2.06	45	0.00				
AC04-08511	040603 2b B4	-0.13	3.21	45	-0.01				
AC04-08512	040603 2b B3	-0.10	33.30	45	0.00				
AC04-08513	040603 2b B2	-0.18	30.69	45	-0.01				
AC04-08514	040603 2b B1	-0.20	19.19	45	-0.01				
AC04-08515	040603 2b C5	-0.24	7.25	45	-0.01				
AC04-08516	040603 2b CP	-0.09	1.89	750	-0.07				
AC04-08517	040603 2b DP	0.01	3.26	1000	0.01				
AC04-08518	040603 2b QL	1.43	4.13	270	0.39	0.497			
AC04-08519	040603 2b QFT1	-0.23	46.04	133	-0.03				
AC04-08520	040603 2b QFT2	-0.03	28.76	78	0.00				
AC04-08521	040603 2b QW	-0.01	9.12	48	0.00				

FIG. 17C

FIG. 18A

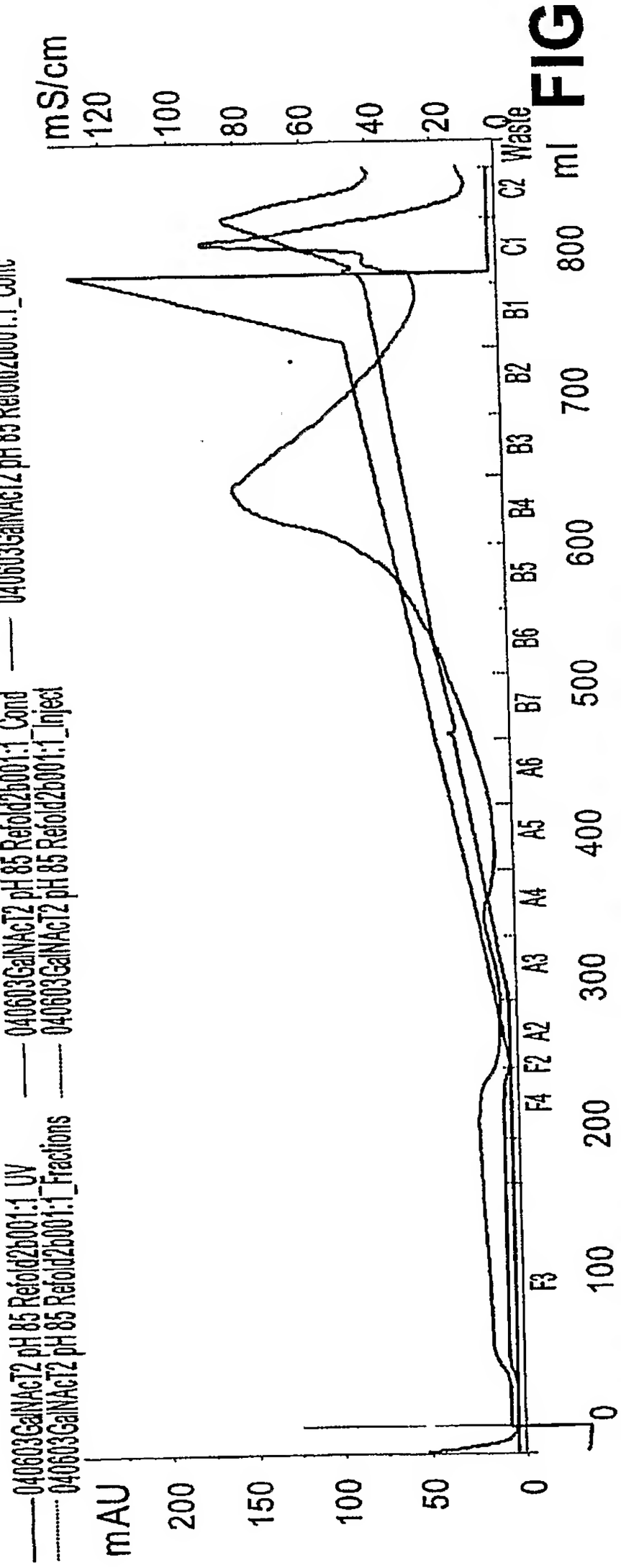
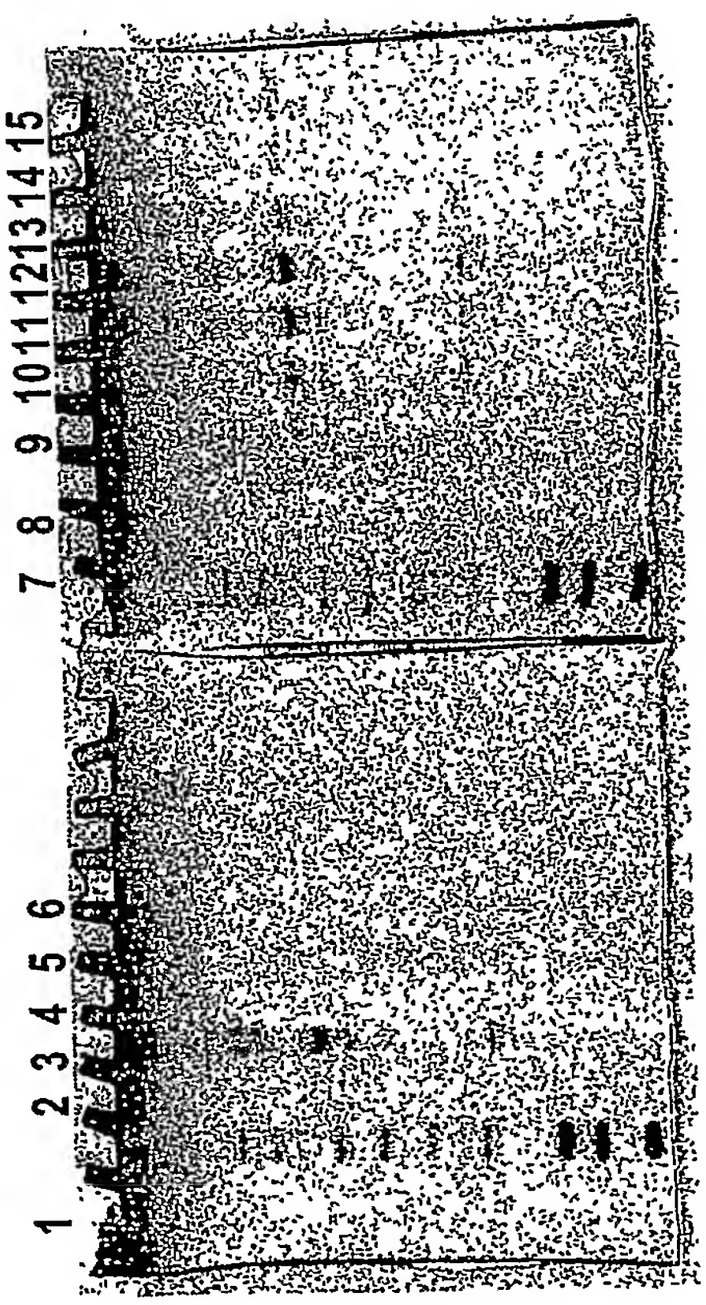


FIG. 18B

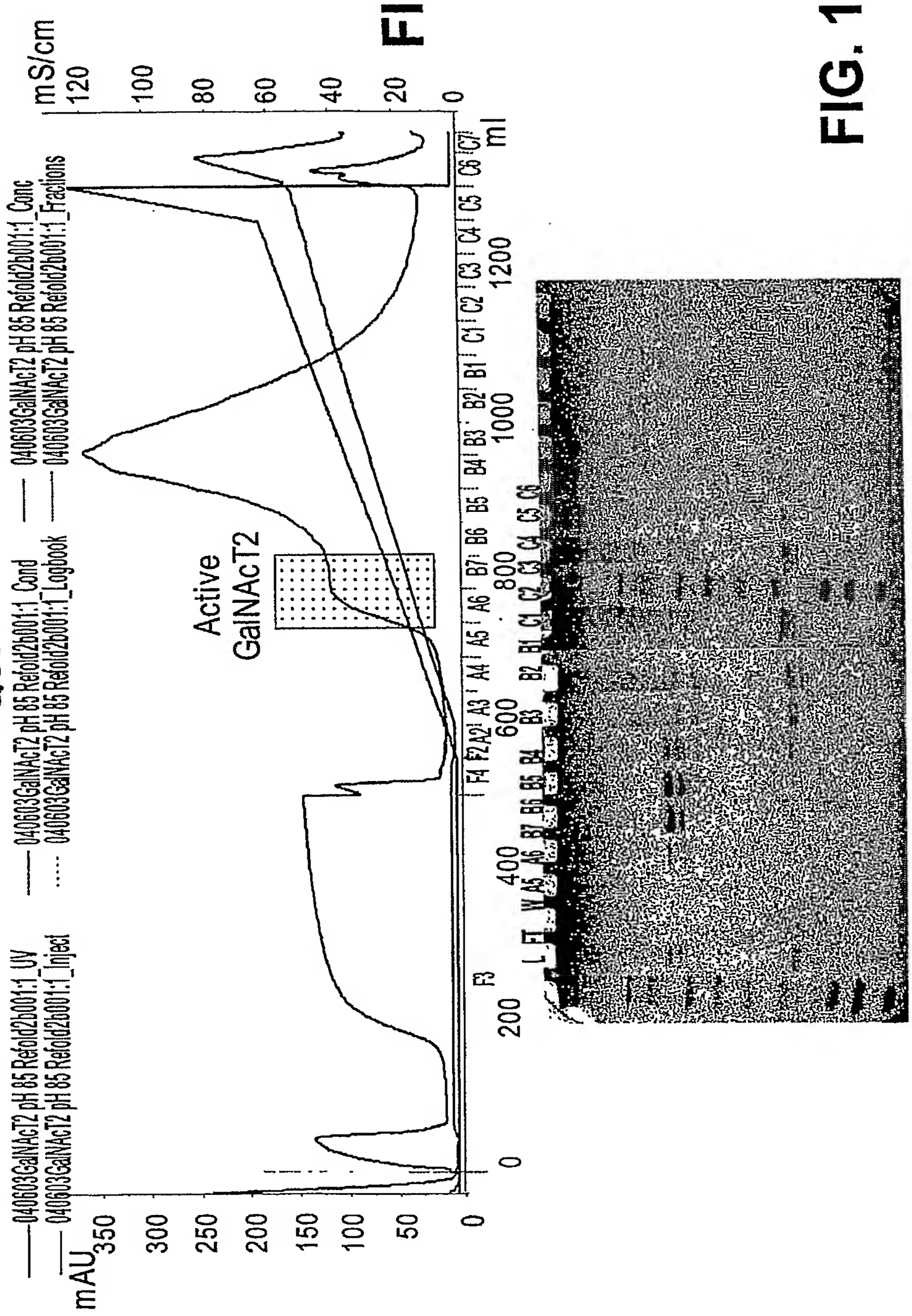


AC#	Sample Description	Activity (U/L)	%RSD	Volume (ml)	Activity (U)
AC04-08522	040603 3 A6	0.14	14.4	45	0.01
AC04-08523	040603 3 B7	0.13	0.0	45	0.01
AC04-08524	040603 3 B6	0.01	1.9	45	0.00
AC04-08525	040603 3 B5	0.02	29.8	45	0.00
AC04-08526	040603 3 B4	-0.05	0.0	45	0.00
AC04-08527	040603 3 B3	-0.03	7.4	45	0.00
AC04-08528	040603 3 B2	-0.01	2.2	45	0.00
AC04-08529	040603 3 B1	-0.02	33.7	45	0.00
AC04-08530	040603 3 C1	-0.03	21.6	45	0.00
AC04-08531	040603 3 CP	-0.02	7.0	750	-0.02
AC04-08532	040603 3 DP	-0.05	8.3	1000	-0.05
AC04-08533	040603 3 QL	0.05	19.7	250	0.01
AC04-08534	040603 3 QFT1	-0.07	3.1	139	-0.01
AC04-08535	040603 3 QFT2	-0.07	18.4	56	0.00
AC04-08536	040603 3 QW	0.04	47.1	48	0.00

FIG. 18C

FIG. 19A

FIG. 19B



	Volume (mL)	Activity (U/L)	A280	A280/ 1.51 (mg/ mL)	Activity (U)	Mass (mg)	Specific Activity (U/mg)
Load	576	2.73	0.581	0.385	1.57	222	0.007
FT	576	0.02	0.155	0.103	0.01	59	
Wash	48	0.00	0.000		0	0	
A5	45		0.016	0.010	0	0.5	
A6	45	7.74	0.162	0.107	0.35	4.8	0.073
B7	45	10.90*	0.218	0.144	0.49	6.5	0.075
B6	45	1.15*	0.249	0.165	0.05	7.4	0.007
B5 -B1	225	0.00	0.482	0.319	0	108	

* Adjusted by multiplication with 2.5 necessary due to 3H-label change

FIG. 19C

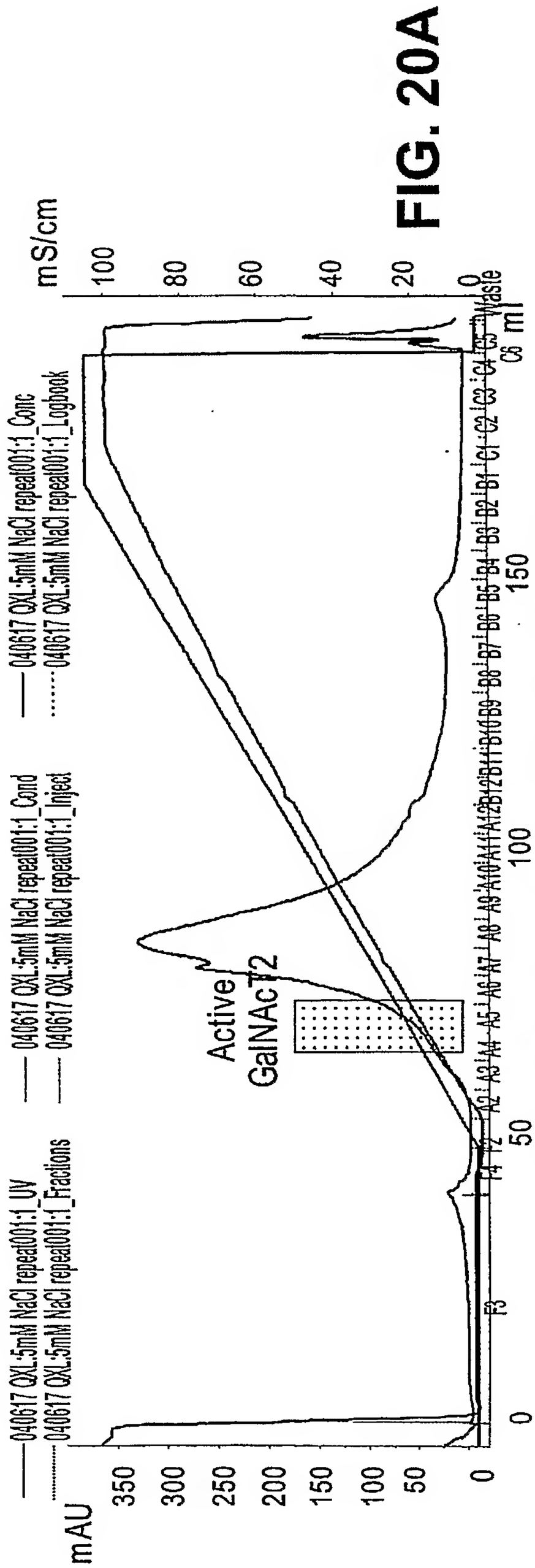


FIG. 20A

FIG. 20B

	Volume (mL)	Activity (U/L)	A280	A280/ 1.51 (mg/mL)	Activity (mU)	Mass (mg)	Specific Activity (U/mg)
Load	40	1.23	0.550	0.364	49.2	14.6	0.003
FT	40	-0.02	-0.017				
Wash	7.8	-0.01	-0.008				
A3	5	0.41	0.000				
A4	5	1.82	0.048	0.032	9.1	0.16	0.057
A5	5	2.14	0.100	0.066	10.7	0.33	0.032
A6-A9	20	1.12	0.422	0.279	22.4	5.58	0.004

FIG. 20C

FIG. 21A

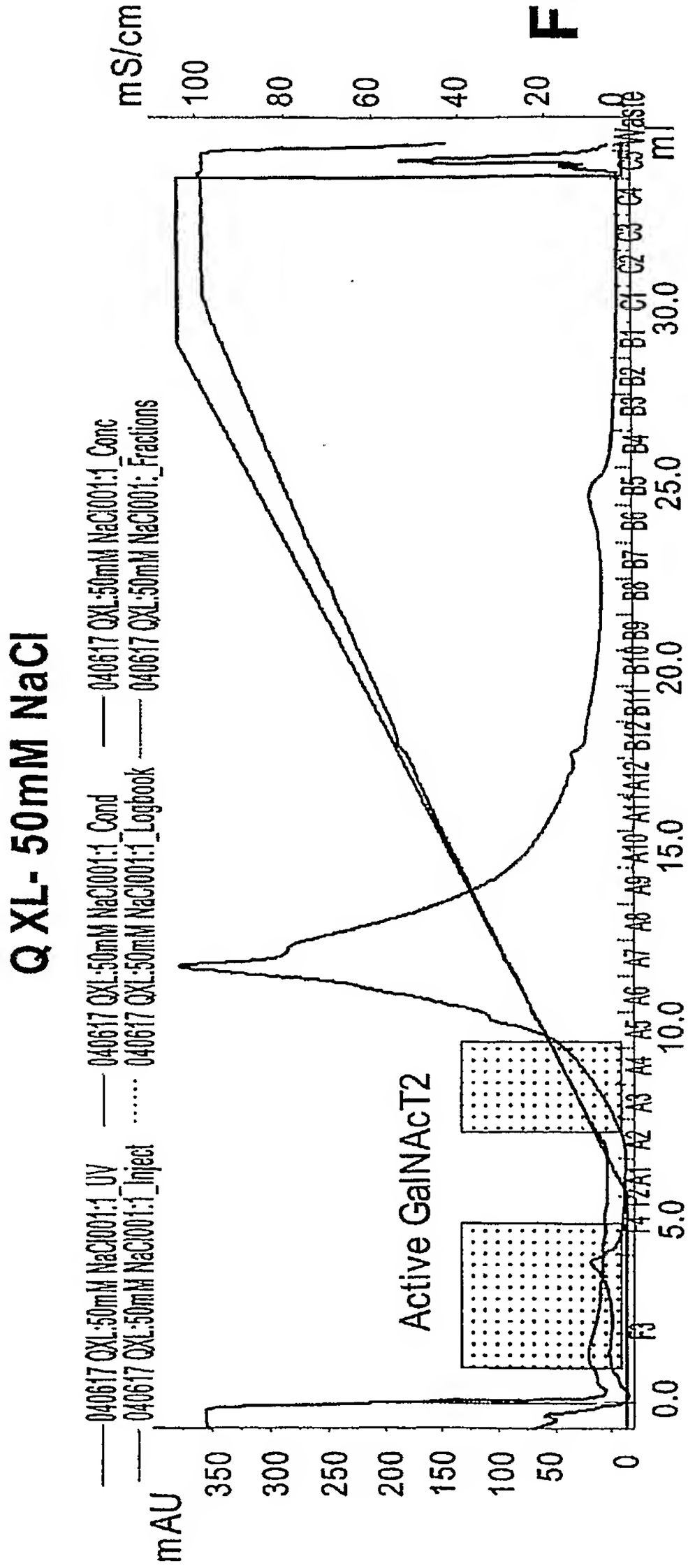
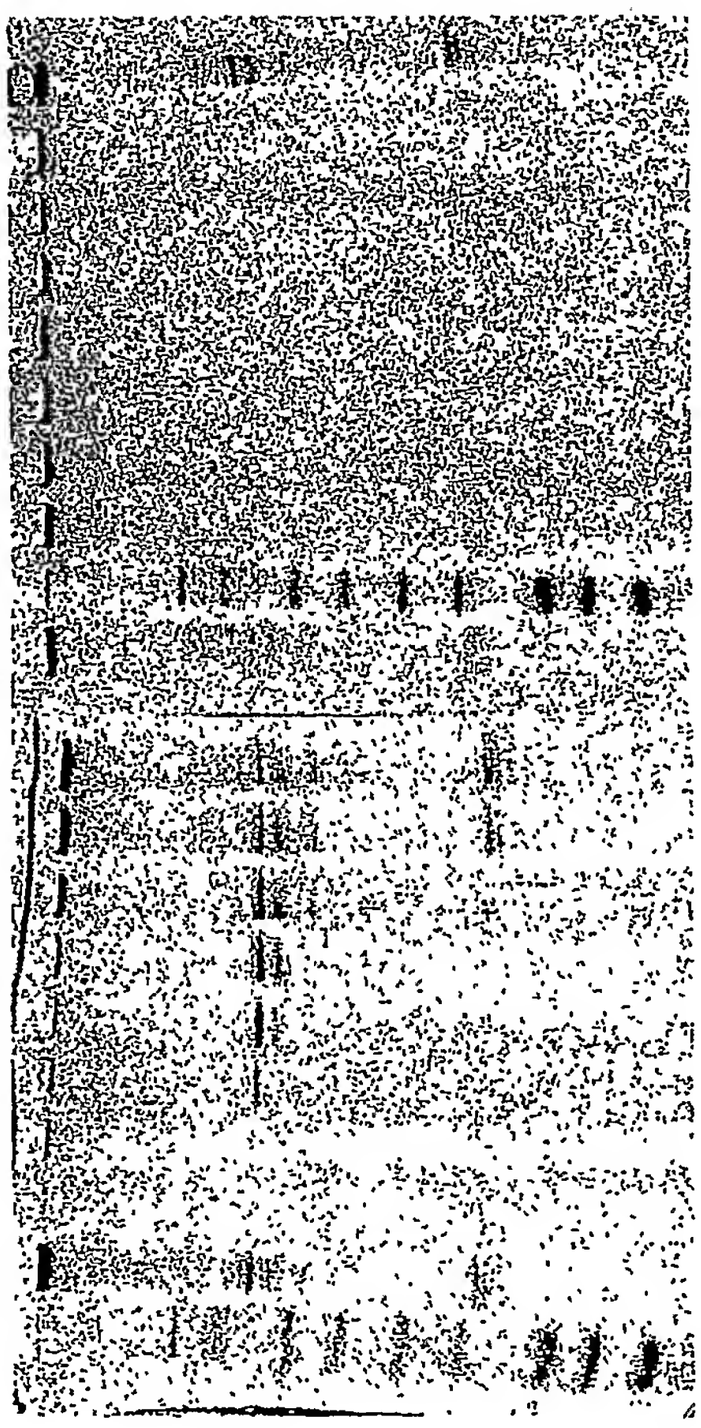


FIG. 21B



	Volume (mL)	Activity (U/L)	A280	A280/ 1.51 (mg/mL)	Activity (mU)	Mass (mg)	Specific Activity (U/mg)
Load	40	1.27	0.550	0.364	50.8	14.6	0.0035
FT	40	0.41	0.001	0.007	16.4	0.03	0.55
Wash	7.8	0.07	-0.007				
A2	5	0.40	-0.018				
A3	5	0.50	0.004	0.003	2.5	0.02	0.125
A4	5	0.87	0.049	0.032	4.4	0.16	0.0275
A5	5		0.136	0.090		0.45	
A6-A9	20	0.29	0.418	0.277	5.8	5.54	0.001

FIG. 21C

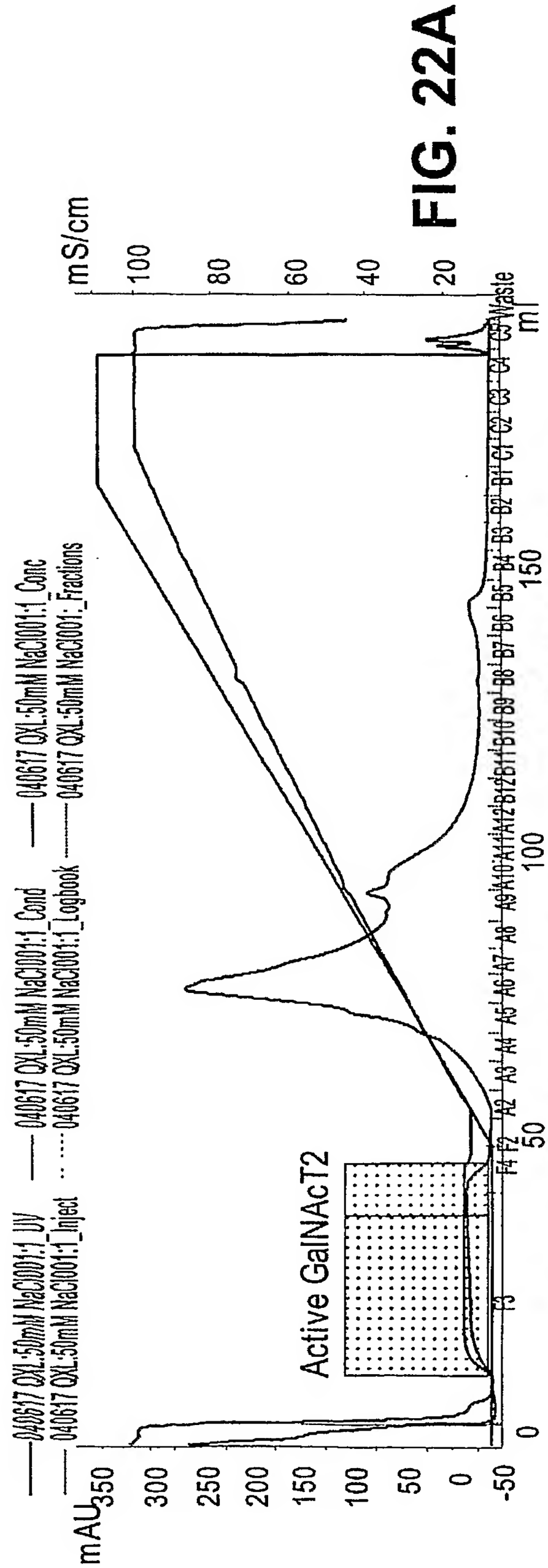


FIG. 22A

FIG. 22B

	Volume (mL)	Activity (U/L)	A280	A280/ 1.51 (mg/mL)	Activity (mU)	Mass (mg)	Specific Activity (U/mg)
Load	40	1.25	0.561	0.372	50.0	14.9	0.0034
FT	40	0.71	0.019	0.012	28.4	0.48	0.059
Wash	7.8	0.48	0.026	0.017	3.7	0.13	0.028
A3	5	0.13	0.011	0.007	0.6	0.04	0.015
A4	5	0.11	0.062	0.041	0.6	0.21	0.003
A5 -A9	25	0.04	0.311	0.206	1.0	5.15	0.0002

FIG. 22C

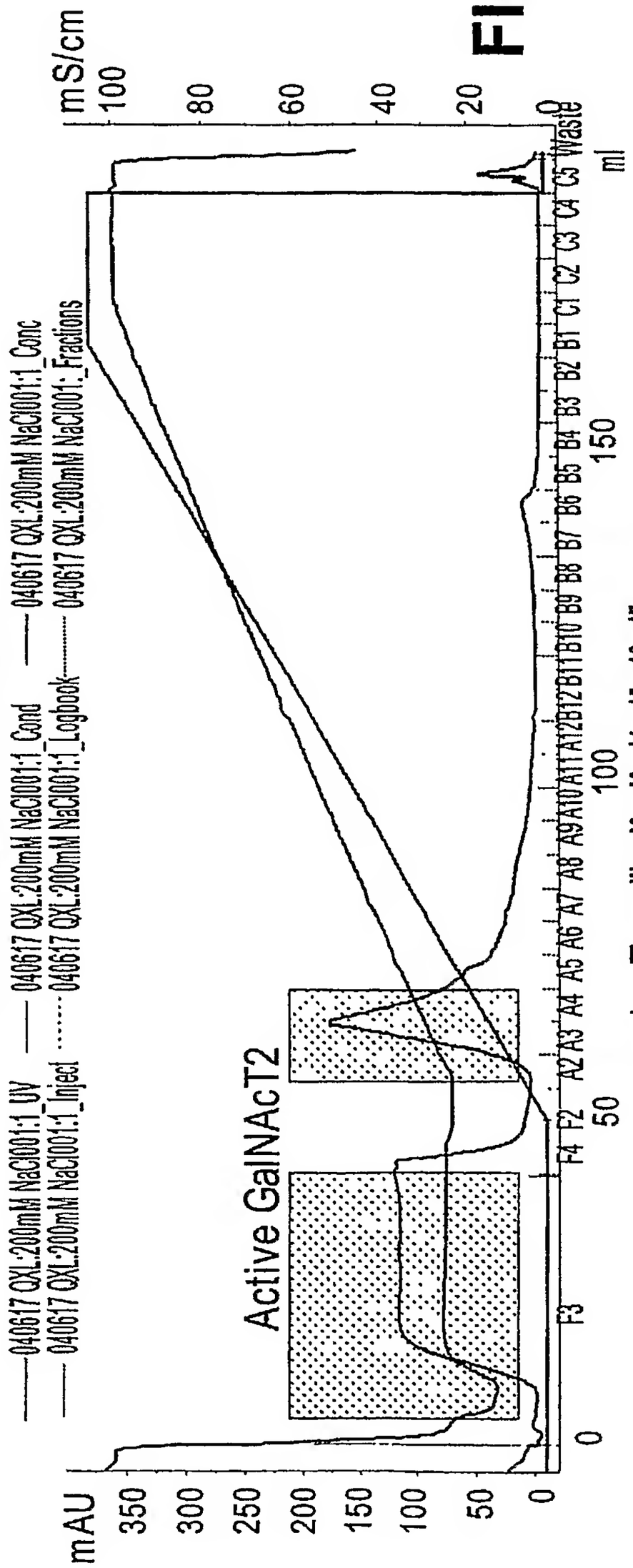


FIG. 23A

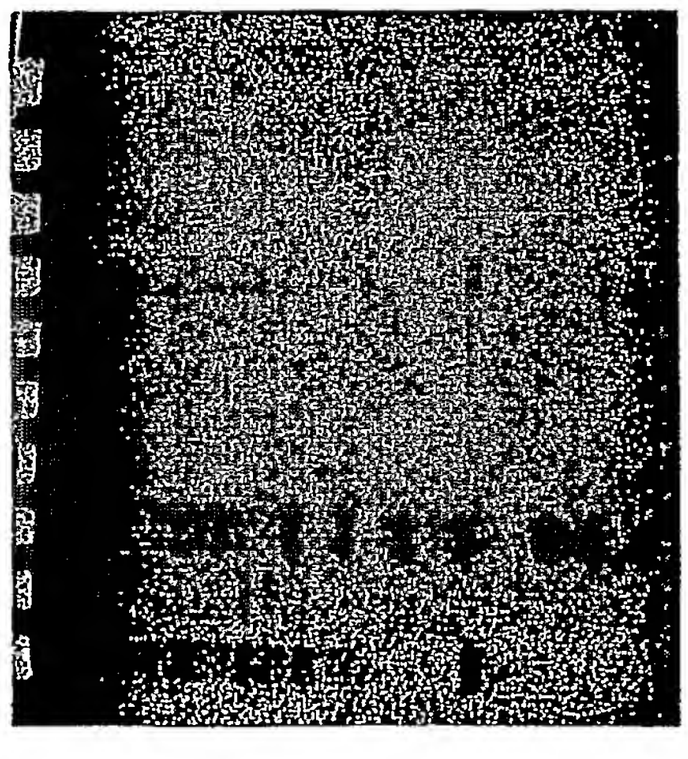


FIG. 23B

	Volume (mL)	Activity (U/L)	A280	A280/ 1.51 (mg/ mL)	Activity (mU)	Mass (mg)	Specific Activity (U/mg)
Load	40	1.20	0.579	0.383	48.0	15.3	0.0031
FT	40	0.81	0.151	0.100	32.4	4.00	0.0081
Wash	7.8	0.01	0.128	0.085	0	0.66	
A3	5	1.65	0.158	0.105	8.25	0.53	0.0156
A4	5	3.76	0.273	0.181	18.8	0.91	0.0206
A5/6	10	0.01	0.093	0.062	0.1	0.62	0.0002

FIG. 23C

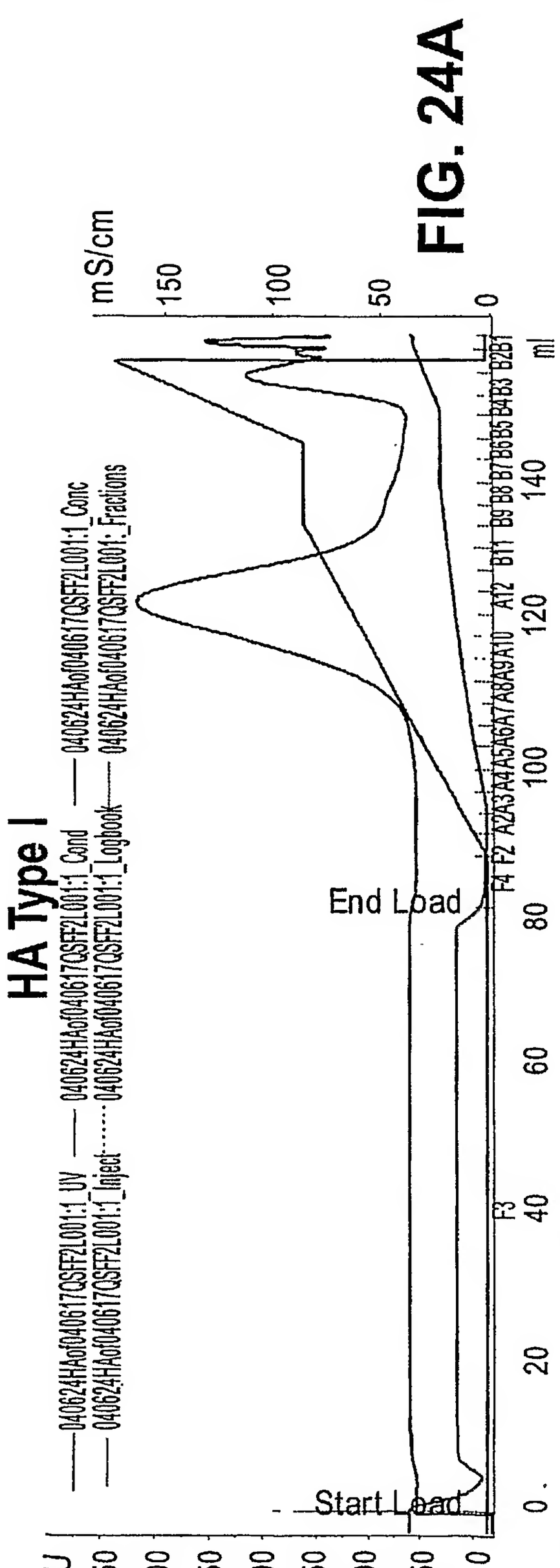


FIG. 24A

	Volume (mL)	Activity (U/L)	Activity (U)
Load	71	7.2	0.511
FT	81	0.0	0
Wash	6.75	0.0	0
A8/9	6	1.0	0.006
A10/11	6	18.3	0.110
A12/B12	6	30.5	0.183
B11/10	6	11.2	0.067
B9	3	0.5	0.002
B3	3	0.1	0

FIG. 24C

FIG. 24B

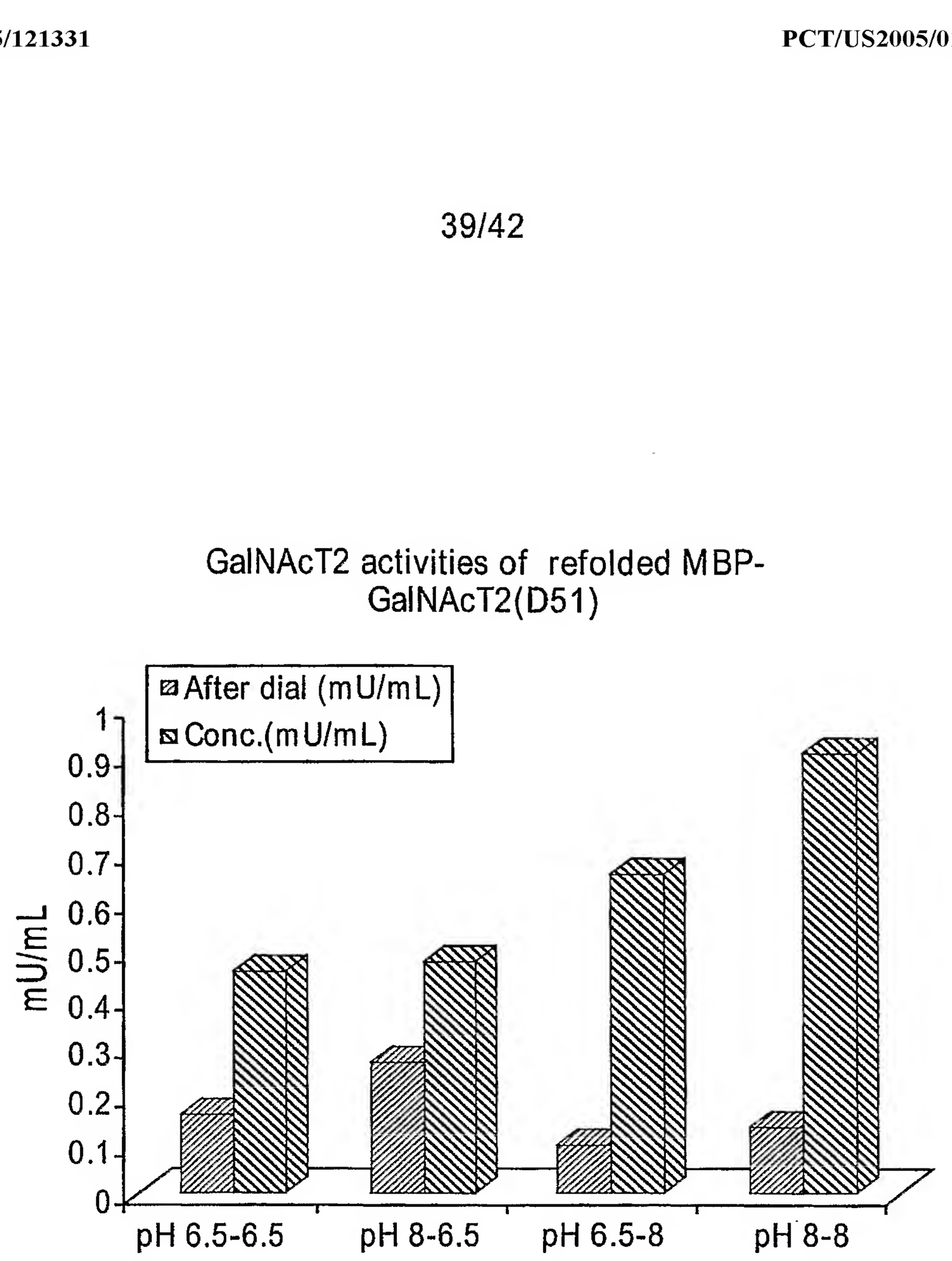

GalNAcT2 activities of refolded MBP-GalNAcT2(D51)
After dial (mU/mL)
Conc.(mU/mL)
mU/mL
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
pH 6.5-6.5
pH 8-6.5
pH 6.5-8
pH 8-8

FIG. 25

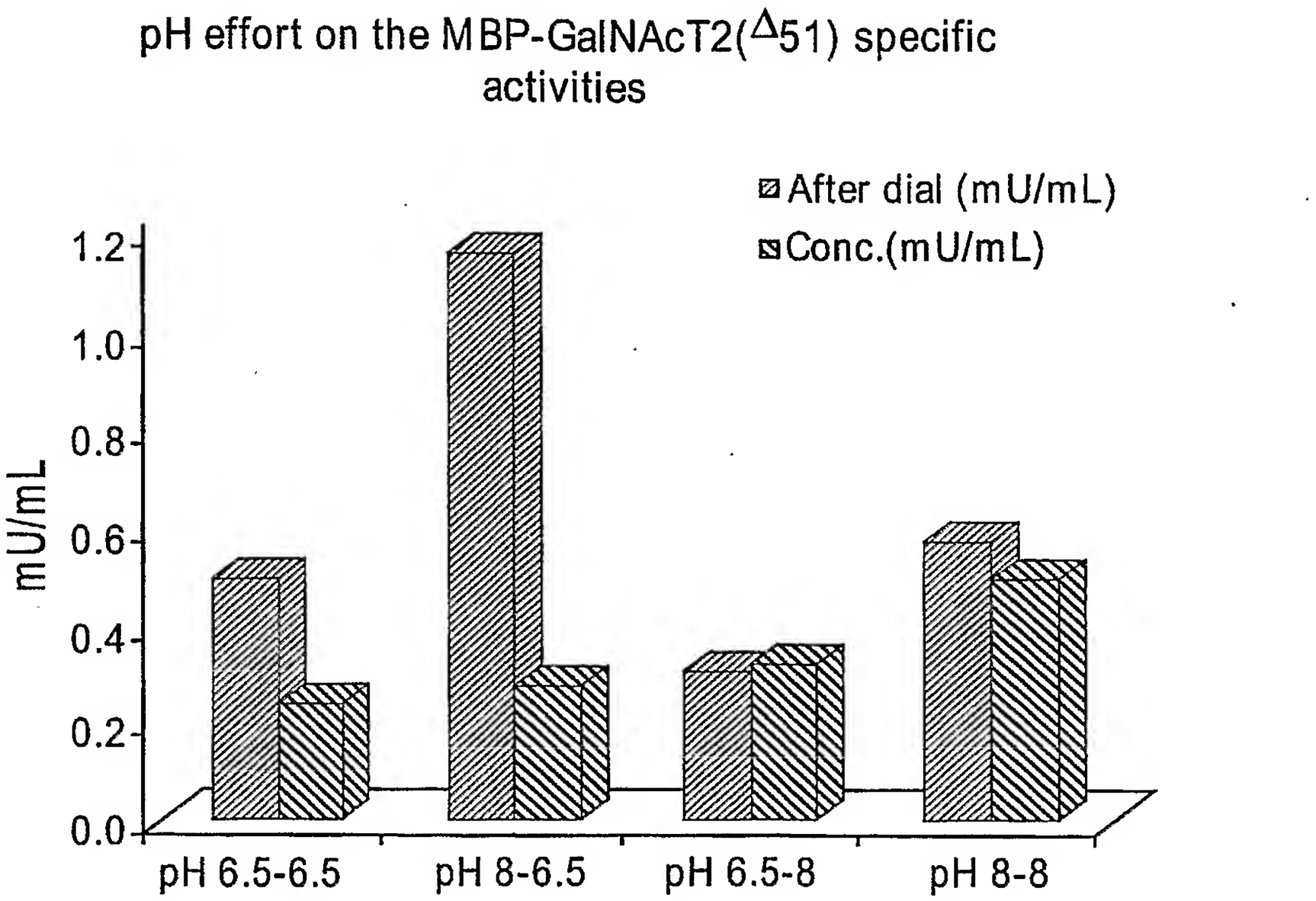

pH effort on the MBP-GalNAcT2(Δ51) specific activities
After dial (mU/mL)
Conc.(mU/mL)
mU/mL
1.2
1.0
0.8
0.6
0.4
0.2
0.0
pH 6.5-6.5
pH 8-6.5
pH 6.5-8
pH 8-8
pH Solubiliz. - refold

FIG. 26

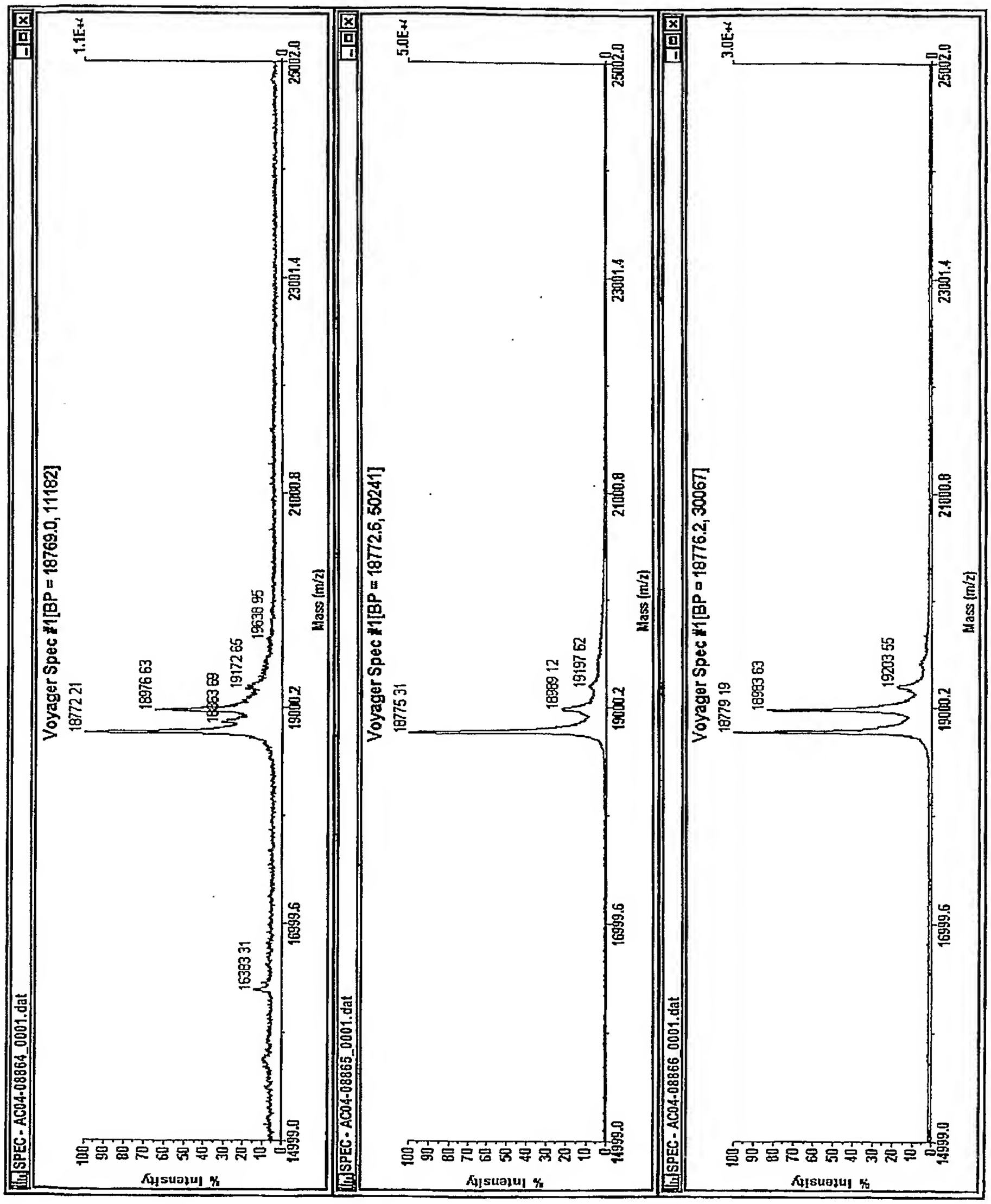

Voyager Spec #1[BP = 18769.0, 11182]
18772.21
18976.63
19172.65
19638.95
Mass (m/z)
SPEC - AC04-08864_0001.dat
% Intensity
16383.31
Voyager Spec #1[BP = 18772.5, 50241]
18775.31
18969.12
19197.62
SPEC - AC04-08865_0001.dat
Voyager Spec #1[BP = 18776.2, 30067]
18779.19
18983.63
19203.55
SPEC - AC04-08866_0001.dat

FIG. 27

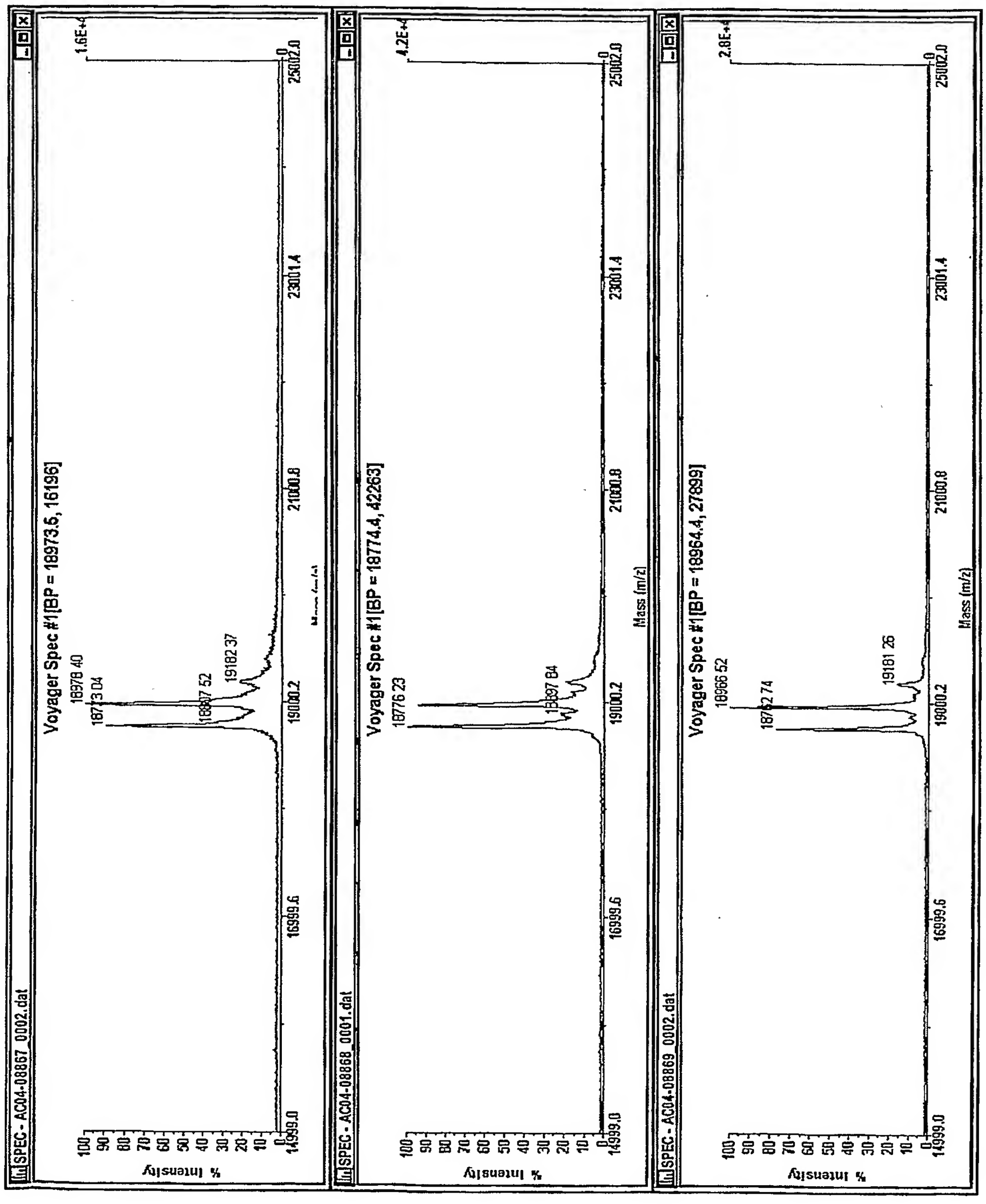

Voyager Spec #1[BP = 18973.5, 16196]
Voyager Spec #1[BP = 18774.4, 42263]
Voyager Spec #1[BP = 18964.4, 27899]
Mass (m/z)
% Intensity
SPEC - AC04-08867_0002.dat
SPEC - AC04-08868_0001.dat
SPEC - AC04-08869_0002.dat

FIG. 28